

GenCore version 5.1.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:48:41 ; Search time 14.8101 Seconds  
(without alignments)  
84.457 Million cell updates/sec

Title: US-10-617-568-36

Perfect score: 44

Sequence: 1 AAXAAXAAXAAXAA 13

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR\_80:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	85	1 FDFL4W	antifreeze protein
2	44	100.0	91	2 A22592	antifreeze protein
3	44	100.0	150	2 T12547	hypothetical prote
4	44	100.0	220	2 JC5954	ribosomal protein
5	44	100.0	233	2 S11563	probable MASH-2 pr
6	44	100.0	238	2 A48279	achaete scute prot
7	44	100.0	289	2 A43562	homeotic protein H
8	44	100.0	305	2 I57039	genomic screen hom
9	44	100.0	314	2 JC5273	paired type homeob
10	44	100.0	323	2 S16318	homeotic protein H
11	44	100.0	331	2 B47236	zinc-finger protei
12	44	100.0	333	2 A39065	homeotic protein E
13	44	100.0	334	2 G02409	protein kinase C-b
14	44	100.0	364	2 I48188	gene Nfx6.1 protel
15	44	100.0	374	2 T03875	probable homeobox
16	44	100.0	375	2 T03874	probable homeobox
17	44	100.0	378	2 A44423	basic helix-loop-h
18	44	100.0	401	2 A48423	engrailed homeodom
19	44	100.0	403	2 A53662	homeotic protein H
20	44	100.0	420	2 T14911	bZIP DNA-binding p
21	44	100.0	443	1 I38239	transcription fact
22	44	100.0	475	2 A43915	homeotic protein e
23	44	100.0	477	2 A47236	zinc-finger protei
24	44	100.0	482	2 JC7583	basic helix-loop-h
25	44	100.0	490	2 T09084	phosphatidylinosit
26	44	100.0	494	2 A42170	zinc finger protei
27	44	100.0	495	1 S11223	transcription fact
28	44	100.0	497	2 JC5076	myc-associated zin
29	44	100.0	513	2 A48233	polyomavirus enhan

30	44	100.0	552	1 WJFPEN	homeotic protein e
31	44	100.0	604	2 A39369	homeotic protein B
32	44	100.0	606	2 S13367	On(1D) protein - f
33	44	100.0	627	2 T02610	probable YME1 ATP-
34	44	100.0	644	2 S39356	transcription fact
35	44	100.0	779	1 S40382	box A-binding fact
36	44	100.0	796	2 JC7555	C14orf4 protein -
37	44	100.0	805	2 T49385	hypothetical prote
38	44	100.0	806	2 T13690	hypothetical prote
39	44	100.0	828	2 C88402	protein H05C05.1 f
40	44	100.0	829	2 A34692	ecdysone-induced p
41	44	100.0	873	2 B53225	ecdysone-induced p
42	44	100.0	883	2 S04722	puff 74E protein -
43	44	100.0	1028	2 A56038	DNA-binding protei
44	44	100.0	1052	2 T14343	zinc finger RNA bi
45	44	100.0	1065	2 T13230	dachshund isoform
46	44	100.0	1072	2 T13232	dachshund protein
47	44	100.0	1074	2 T13229	dachshund protein
48	44	100.0	1081	2 T13231	dachshund protein
49	44	100.0	1180	2 S69205	stripe a/b protein
50	44	100.0	1205	2 A55015	bumetanide-sensiti
51	44	100.0	1212	2 A57187	bumetanide-sensiti
52	44	100.0	1213	2 S16356	ovo protein - frui
53	44	100.0	1355	2 S40022	spalt protein - fr
54	44	100.0	1506	2 T51900	related to PAN2 pr
55	44	100.0	1668	2 T13748	sex comb protein -
56	44	100.0	1891	2 T13594	hypothetical prote
57	44	100.0	2038	2 A43742	female sterile hom
58	44	100.0	2639	2 T31328	fibroin - Chinese
59	44	100.0	5327	2 T13564	microtubule-associ
60	43	97.7	273	2 T51010	hypothetical prote
61	43	97.7	306	1 S59863	polyA binding prot
62	43	97.7	377	2 T04213	heat shock transcr
63	43	97.7	392	2 B48423	homeotic protein e
64	43	97.7	451	2 D88395	protein F53A3.6 [i
65	43	97.7	530	2 I38558	Mi-2 autoantigen 2
66	43	97.7	577	2 S72227	finger protein sob
67	43	97.7	792	1 EAHU	elastin precursor,
68	43	97.7	892	2 T09071	SH3 domains-contai
69	43	97.7	1533	2 A46221	abdominal segment
70	42	95.5	172	2 S35568	sex-determining pr
71	41	93.2	448	2 S17370	DNA-binding protei
72	41	93.2	509	2 T02940	O-succinylhomoseri
73	41	93.2	509	2 T02942	O-succinylhomoseri
74	41	93.2	640	2 A41726	homeotic protein B
75	41	93.2	642	2 S27806	homeotic protein B
76	41	93.2	832	2 A32446	neurogenic protein
77	41	93.2	833	2 S19087	gene Delta protein
78	41	93.2	868	2 A30817	serendipity (sry h
79	41	93.2	880	2 S00670	neurogenic repetit
80	41	93.2	2715	2 T13049	eyelid - fruit fly
81	41	93.2	2957	2 T33152	hypothetical prote
82	41	93.2	3828	2 T13857	trithorax protein
83	41	93.2	4845	2 T31067	BIR repeat contain
84	40	90.9	40	2 S58853	homeotic protein u
85	40	90.9	97	2 S02376	antifreeze protein
86	40	90.9	231	2 S28186	achaete-scute locu
87	40	90.9	376	2 A26066	segmentation prote
88	40	90.9	425	1 JH0710	transcription fact
89	40	90.9	425	2 A38153	paired box protein
90	40	90.9	509	2 A36392	segmentation prote
91	40	90.9	578	2 T22888	hypothetical prote
92	40	90.9	592	2 T22885	hypothetical prote
93	40	90.9	629	2 T22890	hypothetical prote
94	40	90.9	649	2 S43229	arylsulfatase (EC
95	40	90.9	653	2 A39922	potassium channel
96	40	90.9	660	2 S24125	probable Poly-A Bi
97	40	90.9	671	2 C96534	homeotic protein o
98	40	90.9	701	2 A35912	transcription fact
99	40	90.9	702	1 S46458	T-box protein 2 -
100	40	90.9	702	2 G01840	

## ALIGNMENTS

## RESULT 1

FDL4W  
antifreeze protein 4 precursor - winter flounder  
C:Species: Pseudopleuronectes americanus (winter flounder)  
C>Date: 01-Sep-1981 #sequence\_revision 01-Sep-1981 #text\_change 09-Jul-2004  
C:Accession: A03193  
R:Lin, Y.; Gross, J.K.  
Proc. Natl. Acad. Sci. U.S.A. 78, 2825-2829, 1981  
A:Title: Molecular cloning and characterization of winter flounder antifreeze cDNA.  
A:Reference number: A03193; MUID:81247379; PMID:6265915  
A:Accession: A03193  
A:Molecule type: mRNA  
A:Residues: 1-85 <LIN>  
A:CROSS-references: UNIPROT:P02734; UNIPARC:UPI0000174278  
C:Superfamily: antifreeze protein  
C:Keywords: antifreeze  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-85/Product: antifreeze protein 4 #status predicted <MAT>

Query Match 100.0%; Score 44; DB 1; Length 85;  
Best Local Similarity 84.6%; Pred. No. 7;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 35 AATAAAAAATAA 47

## RESULT 2

A22592  
antifreeze protein IIA7 precursor - winter flounder  
C:Species: Pseudopleuronectes americanus (winter flounder)  
C>Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 09-Jul-2004  
R:Gourlie, B.; Lin, Y.; Price, J.; De Vries, A.L.; Powers, D.; Huang, R.C.C.  
J. Biol. Chem. 259, 14960-14965, 1984  
A:Title: Winter flounder antifreeze proteins: a multigene family.  
A:Reference number: A22592; MUID:85054993; PMID:6548752  
A:Accession: A22592  
A:Molecule type: mRNA  
A:Residues: 1-91 <GOU>  
A:CROSS-references: UNIPROT:P07835; UNIPARC:UPI000012584B; GB:M10148; NID:G213579; PIDN:  
R:Gauthier, S.; Wu, Y.; Davies, P.L.  
Nucleic Acids Res. 18, 5303, 1990  
A:Title: Nucleotide sequence of a variant antifreeze protein gene.  
A:Reference number: S12604; MUID:90384854; PMID:2402466  
A:Accession: S12604  
A:Molecule type: DNA  
A:Residues: 1-53, 'V' 55-91 <GAU>  
A:CROSS-references: UNIPARC:UPI0000001382; EMBL:X53718; NID:G64213; PIDN:CAA37754.1; PID  
C:Genetics:  
A:Insertions: 19/2  
C:Superfamily: antifreeze protein  
C:Keywords: antifreeze; tandem repeat  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-91/Product: antifreeze protein IIA7 #status predicted <MAT>

Query Match 100.0%; Score 44; DB 2; Length 91;  
Best Local Similarity 84.6%; Pred. No. 7.3;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 52 AATAAAAAATAA 64

## RESULT 3

T12547  
hypothetical protein DKFZp586E1621.1 - human (fragment)  
C:Species: Homo sapiens (man)

C>Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T12547  
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z17528  
A:Accession: T12547  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-150 <OTT>  
A:CROSS-references: UNIPROT:Q9Y4M1; UNIPARC:UPI0000071DDD; EMBL:AL080235  
A:Experimental source: adult uterus; clone DKFZp586E1621  
C:Genetics:  
A:Note: DKFZp586E1621.1

Query Match 100.0%; Score 44; DB 2; Length 150;  
Best Local Similarity 84.6%; Pred. No. 11;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 129 AAAAAAAAAXAA 141

## RESULT 4

JC5954  
ribosomal protein L14 - human  
C:Species: Homo sapiens (man)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 02-Sep-2000  
C:Accession: JC5954  
R:Tanaka, M.; Tanaka, T.; Harata, M.; Suzuki, T.; Mitsui, Y.  
Biochem. Biophys. Res. Commun. 243, 531-537, 1998  
A:Title: Triplet repeat-containing ribosomal protein L14 gene in immortalized human endo  
A:Reference number: JC5954; MUID:98153799; PMID:9480843  
A:Accession: JC5954  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-220 <TAN>  
A:CROSS-references: UNIPARC:UPI000004E622; DDBJ:DB7735; NID:G1620021; PIDN:BAA13443.1;  
C:Superfamily: rat ribosomal protein L14

Query Match 100.0%; Score 44; DB 2; Length 220;  
Best Local Similarity 84.6%; Pred. No. 14;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 150 AAAAAAAAAXAA 162

## RESULT 5

S11563  
probable MASH-2 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S11563  
R:Johnson, J.E.; Birren, S.J.; Anderson, D.J.  
Nature 346, 858-861, 1990  
A:Title: Two rat homologues of Drosophila achaete-scute specifically expressed in neuro  
A:Reference number: S11562; MUID:90363294; PMID:2392153  
A:Accession: S11563  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-233 <JOH>  
A:CROSS-references: UNIPROT:P19359; UNIPARC:UPI0000047567; EMBL:X53725; NID:G56630; PID  
C:Keywords: DNA binding

Query Match 100.0%; Score 44; DB 2; Length 233;  
Best Local Similarity 84.6%; Pred. No. 15;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 32 AAAAAAAAAXAA 44

## RESULT 6

A48279  
achaete scute protein - human  
C/Species: Homo sapiens (man)  
C/Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 09-Jul-2004  
C/Accession: A48279  
R/Ball, D.W.; Azzoli, C.G.; Baylin, S.B.; Chi, D.; Dou, S.; Donis-Keller, H.; Cumarswami, P.; et al.  
Proc. Natl. Acad. Sci. U.S.A. 90, 5648-5652, 1993  
A/Title: Identification of a human achaete-scute homolog highly expressed in neuroendocrine tissues  
A/Reference number: A48279; MUID:93296195; PMID:8390674  
A/Accession: A48279  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-238 <RES>  
A/Cross-references: UNIPROT:P50553; UNIPARC:UPI0000047566; GB:L08424; NID:G306459; PID:G306459

Query Match 100.0%; Score 44; DB 2; Length 238;  
Best Local Similarity 84.6%; Pred. No. 15;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAXAAXAA 13  
||| ||||| |||

DB 35 AAAAAAAAAAAAAA 47

## RESULT 7

A43562  
homeotic protein Hox D8 - mouse  
N/Alternate names: homeotic protein Hox 4.3  
C/Species: Mus musculus (house mouse)  
C/Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 09-Jul-2004  
C/Accession: A43562  
R/Izpisua-Belmonte, J.C.; Dolle, P.; Renucci, A.; Zappavigna, V.; Falkenstein, H.; Duboulet, J.H.; et al.  
Development 110, 733-745, 1990  
A/Title: Primary structure and embryonic expression pattern of the mouse Hox-4.3 homeobox gene  
A/Reference number: A43562; MUID:91209232; PMID:1982431  
A/Accession: A43562  
A/Status: preliminary; not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 1-289 <12P>  
A/Cross-references: UNIPROT:P23463; UNIPARC:UPI0000177DA6  
C/Superfamily: homeotic protein Hox A7; homeobox homology  
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:196-252/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 289;  
Best Local Similarity 84.6%; Pred. No. 18;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAXAAXAA 13  
||| ||||| |||

DB 15 AAAAAAAAAAAAAA 27

## RESULT 8

I57039  
genomic screen homeobox protein 2 - mouse  
C/Species: Mus sp. (mouse)  
C/Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 05-Oct-2004  
C/Accession: I57039  
R/Hsieh-Li, H.M.; Witte, D.P.; Szucsik, J.C.; Weinstein, M.; Li, H.; Potter, S.S.  
Mech. Dev. 50, 177-186, 1995  
A/Title: Gsh-2, a murine homeobox gene expressed in the developing brain.  
A/Reference number: I57039; MUID:95344993; PMID:7619729  
A/Accession: I57039  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-305 <RES>  
A/Cross-references: UNIPARC:UPI0000024E69; GB:S79041; NID:g1042008; PIDN:AA834947.1; PIDN:AA834947.1  
C/Genetics: Gsh-2

C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:204-260/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 305;  
Best Local Similarity 84.6%; Pred. No. 18;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAXAAXAA 13  
||| ||||| |||

DB 147 AAAAAAAAAAAAAA 159

## RESULT 9

JC5273  
paired type homeobox protein, NBP - human  
C/Species: Homo sapiens (man)  
C/Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 05-Oct-2004  
C/Accession: JC5273  
R/Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.  
DNA Res. 3, 311-320, 1996  
A/Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific homeobox genes  
A/Reference number: JC5272; MUID:97191543; PMID:9039501  
A/Contents: neuroblastoma cell  
A/Accession: JC5273  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-314 <YOK>  
A/Cross-references: UNIPROT:Q9453; UNIPARC:UPI0000131D16; DBJ:D82344; NID:g1841337; PIDN:G1841337  
C/Comment: This protein is a transcriptional repressor involved in regulating gene expression  
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:99-155/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 314;  
Best Local Similarity 84.6%; Pred. No. 19;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAXAAXAA 13  
||| ||||| |||

DB 241 AAAAAAAAAAAAAA 253

## RESULT 10

S16318  
homeotic protein Hox 4.6 - mouse  
N/Alternate names: homeotic protein Hoxd-11  
C/Species: Mus musculus (house mouse)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Oct-2004  
C/Accession: S16318; S40403; S57443  
R/Izpisua-Belmonte, J.C.; Falkenstein, H.; Dolle, P.; Renucci, A.; Duboule, D.; et al.  
EMBO J. 10, 2279-2289, 1991  
A/Title: Murine genes related to the Drosophila AbdB homeotic gene are sequentially expressed in the developing mouse embryo  
A/Reference number: S16317; MUID:91293104; PMID:1676674  
A/Accession: S16318  
A/Molecule type: DNA  
A/Residues: 1-323 <12P>  
A/Cross-references: UNIPROT:P23813; UNIPARC:UPI0000029F10; EMBL:X60761; NID:g51420; PIDN:G51420  
R/Gerard, M.; Duboule, D.; Zakany, J.  
EMBO J. 12, 3539-3550, 1993  
A/Title: Structure and activity of regulatory elements involved in the activation of the Hox-4.6 homeobox gene  
A/Reference number: S40403; MUID:94074553; PMID:7902810  
A/Accession: S40403  
A/Status: translation not shown  
A/Molecule type: DNA  
A/Residues: 1-323 <GER>  
A/Cross-references: UNIPARC:UPI0000029F10; EMBL:X71422; NID:g397508; PIDN:CAAS0553.1; PIDN:CAAS0553.1  
R/Takada, S.; Cook, M.; Kramlauf, R.; McMahon, A.P.  
submitted to the EMBL Data Library, May 1991  
A/Description: Genomic sequence of mouse Hox-4.6.  
A/Reference number: S57443  
A/Accession: S57443  
A/Molecule type: DNA  
A/Residues: 'MNDFDCGSAAS', 1-323 <TAK>  
A/Cross-references: UNIPARC:UPI0000026A38; EMBL:X60395; NID:g871427; PIDN:CAA42943.1; PIDN:CAA42943.1

## C;Genetics:

A;Gene: Hoxd-11  
A;Introns: 246/1  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;252-308/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 323;  
Best Local Similarity 84.6%; Pred. No. 19;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
Db 93 AAAAAAAXAAAAA 105

## RESULT 11

B47236  
zinc-finger protein Pur-1 - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C;Accession: B47236  
R;Kennedy, G.C.; Rutter, W.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992  
A;Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates transcription  
A;Reference number: A47236; MUID:93087555; PMID:1454839  
A;Accession: B47236  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-331 <KEN>  
A;Cross-references: UNIPARC:UPI000017C615  
A;Experimental source: insulinoma cell line T  
A;Note: sequence extracted from NCBI backbone (NCBIP:119832)

Query Match 100.0%; Score 44; DB 2; Length 331;  
Best Local Similarity 84.6%; Pred. No. 20;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
Db 83 AAAAAAAXAAAAA 95

## RESULT 12

A39065  
homeotic protein EVX2 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 31-Dec-2004  
C;Accession: A39065  
R;D'Esposito, M.; Morelli, F.; Acampora, D.; Migliaccio, E.; Simeone, A.; Boncinelli, E.  
Genomics 10, 43-50, 1991  
A;Title: EVX2, a human homeobox gene homologous to the even-skipped segmentation gene, is expressed in the developing mouse embryo  
A;Reference number: A39065; MUID:91257849; PMID:1675198  
A;Accession: A39065  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-333 <DAE>  
A;Cross-references: UNIPARC:UPI000017A29D

Query Match 100.0%; Score 44; DB 2; Length 333;  
Best Local Similarity 84.6%; Pred. No. 20;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
Db 213 AAAAAAAXAAAAA 225

A;Gene: GDB:EVX2  
A;Cross-references: GDB:127528; OMIM:142991  
A;Map position: 2q24.3-2q31  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;46-102/Domain: homeobox homology <HOX>

## RESULT 13

G02409  
protein kinase C-binding protein RACK17 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 05-Nov-1999  
C;Accession: G02409  
R;Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Kikkawa, U.  
submitted to the EMBL Data Library, February 1996  
A;Reference number: H01212  
A;Accession: G02409  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-334 <KUR>  
A;Cross-references: UNIPARC:UPI000016A2D0; EMBL:U48250; NID:g1199656; PIDN:AAC72247.1;

Query Match 100.0%; Score 44; DB 2; Length 334;  
Best Local Similarity 84.6%; Pred. No. 20;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
Db 236 AAAAAAAXAAAAA 248

## RESULT 14

I48188  
gene NKx6.1 protein - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Oct-2004  
C;Accession: I48188  
R;Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.  
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994  
A;Title: Preactic beta cells express a diverse set of homeobox genes.  
A;Reference number: I48185; MUID:95083670; PMID:7991607  
A;Accession: I48188  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-364 <RES>  
A;Cross-references: UNIPROT:Q60554; UNIPARC:UPI00001301FA; EMBL:X81409; NID:g587466; PIDN:

Query Match 100.0%; Score 44; DB 2; Length 364;  
Best Local Similarity 84.6%; Pred. No. 21;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
Db 137 AAAAAAAXAAAAA 149

Query Match 100.0%; Score 44; DB 2; Length 364;  
Best Local Similarity 84.6%; Pred. No. 21;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

## RESULT 15

T03875  
probable homeobox protein OSH45, splice form OSH45 [similarity] - rice  
C;Species: Oryza sativa (rice)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Oct-2004  
C;Accession: T03875  
R;Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Ma  
Plant J. 7, 927-938, 1995  
A;Title: Alternative RNA products from a rice homeobox gene.  
A;Reference number: Z15126; MUID:95322999; PMID:7599652  
A;Accession: T03875  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-374 <TAM>  
A;Cross-references: UNIPROT:P93423; UNIPARC:UPI000000AB610; EMBL:D49704; NID:g1805615; PIDN:

Query Match 100.0%; Score 44; DB 2; Length 333;  
Best Local Similarity 84.6%; Pred. No. 20;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
Db 213 AAAAAAAXAAAAA 225

A;Gene: H45  
A;Introns: 174/3; 215/1; 277/3; 321/3; 367/2



C;Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus  
F;291-352/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 374;  
Best Local Similarity 84.6%; Pred. No. 21;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
DB 64 AAAAAAAXAAAAA 76

## RESULT 16

T03874

probable homeobox protein OSH45, splice form OSH44 [similarity] - rice

N;Contains: probable homeobox protein OSH45, splice form OSH42

C;Species: Oryza sativa (rice)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Oct-2004

C;Accession: T03874; T03876

R;Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Matsuda, T.; Plant J. 7, 927-938, 1995

A;Title: Alternative RNA products from a rice homeobox gene.

A;Reference number: 215126; MUID:95322999; PMID:7599652

A;Accession: T03874

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-375 <TAM1>

A;Cross-references: UNIPROT:P93424; UNIPARC:UPI00000AASD5; EMBL:D49704; NID:g1805615; PIR:197-367

A;Experimental source: cv. Nipponbare; splice form OSH44

A;Accession: T03876

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 197-367 <TAM2>

A;Cross-references: UNIPARC:UPI000016DEBB; EMBL:D49704; NID:g1805615; PIDN:BAA08554.1; PIR:197-367

A;Experimental source: cv. Nipponbare; splice form OSH42

C;Comment: For an alternative splice form, see PIR:T03875.

C;Genetics:

A;Gene: H45

A;Introns: 174/3; 215/1; 277/3; 321/3; 367/2

C;Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus  
F;1-375/Product: probable homeobox protein OSH45, splice form OSH44 #status predicted <M

F;197-367/Product: probable homeobox protein OSH45, splice form OSH42 #status predicted

F;291-352/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 375;  
Best Local Similarity 84.6%; Pred. No. 21;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
DB 64 AAAAAAAXAAAAA 76

## RESULT 17

A44443

basic helix-loop-helix transcription repressor H - fruit fly (Drosophila virilis)

C;Species: Drosophila virilis

C;Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Jul-1998

C;Accession: A44443

R;Wainwright, S.M.; Ish-Horowicz, D.

Mol. Cell. Biol. 12, 2475-2483, 1992

A;Title: Point mutations in the Drosophila hairy gene demonstrate in vivo requirements for DNA binding and transcriptional activation

A;Reference number: A44443; MUID:92269619; PMID:1588951

A;Accession: A44443

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-378 <WAI>

A;Cross-references: UNIPARC:UPI000017BEF1

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIN:102964, NCBI:P:102966)

C;Genetics:

A;Gene: FlyBase:Dvir/h

A;Cross-references: FlyBase:FBgn0013115

Query Match 100.0%; Score 44; DB 2; Length 378;  
Best Local Similarity 84.6%; Pred. No. 22;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
DB 269 AAAAAAAXAAAAA 281

## RESULT 18

A48423

engrailed homeodomain-containing protein En-1 - mouse

N;Alternate names: homeotic protein En-1

C;Species: Mus musculus (house mouse)

C;Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004

C;Accession: A48423; S13009; A26629; A24778

R;Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.

Dev. Genet. 13, 345-358, 1992

A;Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene

A;Reference number: A48423; MUID:93185339; PMID:1363401

A;Accession: A48423

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-401 <LOG>

A;Cross-references: UNIPROT:P09065; UNIPARC:UPI00000299A1

A;Experimental source: CD-1, embryo

A;Note: sequence extracted from NCBI backbone (NCBI:P:126620)

R;Holland, P.W.H.; Williams, N.A.

FEBS Lett. 277, 250-252, 1990

A;Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution.

A;Reference number: S13009; MUID:91099509; PMID:1980115

A;Accession: S13009

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 321-380 <HOL>

A;Cross-references: UNIPARC:UPI000017A2AF

R;Joyner, A.L.; Martin, G.R.

Genes Dev. 1, 29-38, 1987

A;Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrailed gene

A;Reference number: A91620; MUID:88112776; PMID:2892757

A;Accession: A26629

A;Molecule type: DNA; mRNA

A;Residues: 278-401 <JOY>

A;Cross-references: UNIPARC:UPI000016CA7C; GB:Y00201; GB:M11987; NID:g49587; PIDN:CAA68

R;Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.

Cell 43, 29-37, 1985

A;Title: Expression during embryogenesis of a mouse gene with sequence homology to the

A;Reference number: A24778; MUID:86079501; PMID:2416459

A;Accession: A24778

A;Molecule type: DNA

A;Residues: 311-401 <JO2>

A;Cross-references: UNIPARC:UPI000017A2B0

C;Genetics:

A;Gene: en.1

A;Map position: 1

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

F;313-369/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 401;  
Best Local Similarity 84.6%; Pred. No. 23;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
||| ||||| |||  
DB 207 AAAAAAAXAAAAA 219

## RESULT 19

A53662

homeotic protein HB9 - human

C;Species: Homo sapiens (man)

C;Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 31-Dec-2004

C;Accession: A53662  
R;Harrison, K.A.; Druey, K.M.; Deguchi, Y.; Tuscano, J.M.; Kehrl, J.H.  
J. Biol. Chem. 269, 19968-19975, 1994  
A;Title: A novel human homeobox gene distantly related to proboscipedia is expressed in A;Reference number: A53662; MUID:94327547; PMID:7914194  
A;Accession: A53662  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-403 <HAR>  
A;Cross-references: UNIPARC:UPI000017A29E; GB:U07663  
A;Note: the nucleotide sequence and conceptual translation as given are self-consistent  
C;Genetics:  
A;Gene: GDB:HLXB9  
A;Cross-references: GDB:136411; OMIM:142994  
A;Map position: 1q41-1q42.1  
A;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;244-300/Domain: homeobox homology <HOX>  
  
Query Match 100.0%; Score 44; DB 2; Length 403;  
Best Local Similarity 84.6%; Pred. No. 23;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 121 AAAAAAAXAAAAA 133  
  
RESULT 20  
T14911  
BZIP DNA-binding protein - parsley  
C;Species: Petroselinum crispum (parsley)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 31-Dec-2004  
C;Accession: T14911  
R;Kircher, S.; Ledger, S.; Hayashi, H.; Weishaar, B.; Schafer, E.; Frohnmeyer, H.  
Mol. Gen. Genet. 257, 595-605, 1998  
A;Title: CPRF4a, a novel plant bZIP protein of the CPRF family: comparative analysis of A;Reference number: T18261; MUID:98265918; PMID:9604882  
A;Accession: T14911  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-420 <KIR>  
A;Cross-references: UNIPROT:O82037; UNIPARC:UPI00000A12C8; EMBL:Y10810; NID:G3336904; P1  
C;Genetics:  
A;Note: CPRF4b  
C;Superfamily: BZIP G-box-binding protein; fos/jun DNA-binding domain homology  
C;Keywords: DNA binding; leucine zipper  
  
Query Match 100.0%; Score 44; DB 2; Length 420;  
Best Local Similarity 84.6%; Pred. No. 23;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 31 AAAAAAAXAAAAA 43  
  
RESULT 21  
I38239  
transcription factor SOX3 - human  
N;Alternate names: SRY (sex determining region Y)-box 3  
C;Species: Homo sapiens (man)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
A;Accession: I38239; I38242; S67816  
R;Stevanovic, M.; Lovell-Badge, R.; Collignon, J.; Goodfellow, P.N.  
Hum. Mol. Genet. 2, 2013-2018, 1993  
A;Title: SOX3 is an X-linked gene related to SRY.  
A;Reference number: I38239; MUID:94154672; PMID:8111369  
A;Accession: I38239  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-443 <STEL>  
A;Cross-references: UNIPROT:P41225; UNIPARC:UPI000003F546; EMBL:X71135; NID:G468790; PID  
A;Accession: I38242

A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 143-158, 'P', 160-218 <STE2>  
A;Cross-references: UNIPARC:UPI0000071C8C; EMBL:X71137; NID:G468793; PIDN:CAA50467.1; P1  
C;Genetics:  
A;Gene: GDB:SOX3; SOX-3; SOXB  
A;Cross-references: GDB:250376; OMIM:313430  
A;Map position: Xq26-Xq27  
C;Superfamily: human SOX3 protein; HMG box homology  
F;136-211/Domain: HMG box homology <HMG>  
  
Query Match 100.0%; Score 44; DB 1; Length 443;  
Best Local Similarity 84.6%; Pred. No. 24;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 234 AAAAAAAXAAAAA 246  
  
RESULT 22  
A43915  
homeotic protein even-skipped homolog Evx-2 - mouse  
N;Alternate names: mesodermal cell dorsoventral fates determinant Evx-2  
C;Species: Mus musculus (house mouse)  
C;Date: 10-Mar-1999 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C;Accession: A43915  
R;Dush, M.K.; Martin, G.R.  
Dev. Biol. 151, 273-287, 1992  
A;Title: Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive A;Reference number: A43915; MUID:92249649; PMID:1349539  
A;Accession: A43915  
A;Status: preliminary  
A;Molecule type: DNA; mRNA  
A;Residues: 1-475 <DUS>  
A;Cross-references: UNIPROT:P49749; UNIPARC:UPI000020BC6; GB:S34322; NID:G249137; PIDN: A;Note: sequence extracted from NCBI backbone (NCBIN:100478, NCBIN:100480, NCBIN:100482, C;Genetics:  
A;Gene: evx-2  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;192-248/Domain: homeobox homology <HOX>  
  
Query Match 100.0%; Score 44; DB 2; Length 475;  
Best Local Similarity 84.6%; Pred. No. 26;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 359 AAAAAAAXAAAAA 371  
  
RESULT 23  
A47236  
zinc-finger protein Pur-1 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C;Accession: A47236  
R;Kennedy, G.C.; Rutter, W.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992  
A;Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates A;Reference number: A47236; MUID:93087555; PMID:1454839  
A;Accession: A47236  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: nucleic acid  
A;Residues: 1-477 <KENS>  
A;Cross-references: UNIPARC:UPI000017C98F  
A;Experimental source: islet cell line beta TC3  
A;Note: sequence extracted from NCBI backbone (NCBIP:119831)  
  
Query Match 100.0%; Score 44; DB 2; Length 477;  
Best Local Similarity 84.6%; Pred. No. 26;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 96 AAAAAAAAXAA 108

## RESULT 24

JC7583

basic helix-loop-helix protein, DEC2 - human

C:Species: Homo sapiens (man)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004  
C:Accession: JC7583  
R:Fujimoto, K.; Shen, M.; Nishiro, M.; Matsubara, K.; Shingu, S.; Honda, K.; Yoshida, E.  
Biochem. Biophys. Res. Commun. 280, 164-171, 2001  
A:Title: Molecular cloning and characterization of DEC2, a new member of basic helix-loop-helix-loop-helix protein family  
A:Reference number: JC7583; MUID:21092582; PMID:11162494  
A:Accession: JC7583  
A:Molecule type: mRNA  
A:Residues: 1-482 <FUJ>

A:CROSS-references: UNIPROT:Q9COJ9; UNIPARC:UPI0000052B52; DBJ:AB044088  
C:Comment: This protein, a novel member of the DEC subfamily of basic helix-loop-helix protein family  
C:Genetics:  
A:Gene: dec2  
A:Map position: 12p11.23-p12.1  
A:Keywords: transcription factor  
F:1-173/Region: highly conserved #status predicted  
F:130-173/Domain: Orange #status predicted <ORA>  
F:286-411/Region: alanine and glycine-rich #status predicted

Query Match 100.0%; Score 44; DB 2; Length 482;  
Best Local Similarity 84.6%; Pred. No. 26;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 396 AAAAAAAAXAA 408

## RESULT 25

T09084

phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)

C:Species: Chlamydomonas reinhardtii  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T09084  
R:Molendijk, A.J.; Irvine, R.F.  
Plant Mol. Biol. 37, 53-56, 1998  
A:Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinositol 3-kinase  
A:Reference number: Z16411; MUID:98281574; PMID:9620264  
A:Accession: T09084  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-490 <MOL>

A:CROSS-references: UNIPROT:Q04270; UNIPARC:UPI000009C6FA; EMBL:U97663; NID:g2109290; P109084  
A:Experimental source: strain cw-15  
C:Genetics:  
A:Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match 100.0%; Score 44; DB 2; Length 490;  
Best Local Similarity 84.6%; Pred. No. 26;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 115 AATAAAAAAAXAA 127

## RESULT 26

A42170

zinc finger protein MAZ - human (fragment)

N:Alternate names: MYC-associated zinc finger protein MAZ; zinc finger protein ZF87  
C:Species: Homo sapiens (man)  
C:Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 09-Jul-2004  
C:Accession: A42170; A46153  
R:Pyrc, J.J.; Moberg, K.H.; Hall, D.J.

Biochemistry 31, 4102-4110, 1992  
A:Title: Isolation of a novel cDNA encoding a zinc-finger protein that binds to two sites  
A:Reference number: A42170; MUID:92232709; PMID:1567856  
A:Accession: A42170  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-494 <PYR>  
A:CROSS-references: UNIPROT:P56270; UNIPARC:UPI0000017C427; GB:J05371  
A:Note: It is uncertain whether Met-18 is the initiator or whether translation is initiated at Met-19  
R:Bosstone, S.A.; Asselin, C.; Patel, A.J.; Marcu, K.B.  
Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456, 1992  
A:Title: MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcription  
A:Reference number: A46153; MUID:92366479; PMID:1502157  
A:Accession: A46153  
A:Molecule type: mRNA  
A:Residues: 18-417, 'L', 419-494 <BOS>  
A:CROSS-references: UNIPARC:UPI000012ECF8; GB:M94046  
A:Experimental source: HeLa cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:110666, NCBIP:110667)  
C:Keywords: DNA binding; zinc finger  
F:113-125/Region: alanine-rich  
F:174-183/Region: alanine-rich  
F:207-230/Region: zinc finger  
F:296-316/Region: zinc finger  
F:324-346/Region: zinc finger  
F:354-368/Region: zinc finger  
F:373-405/Region: zinc finger  
F:409-430/Region: zinc finger  
F:452-468/Region: alanine-rich

Query Match 100.0%; Score 44; DB 2; Length 494;  
Best Local Similarity 84.6%; Pred. No. 26;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 113 AAAAAAAAXAA 125

## RESULT 27

S31223

transcription factor Brn-1 - mouse

N:Alternate names: class III POU domain protein brain-1  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C:Accession: S31223  
R:Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992  
A:Title: Structure and evolution of four POU domain genes expressed in mouse brain.  
A:Reference number: S31223; MUID:92228768; PMID:1565620  
A:Accession: S31223  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-495 <HAR>  
A:CROSS-references: UNIPROT:P31361; UNIPARC:UPI0000029AAF; EMBL:M88299; NID:g200444; P113123  
C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:27-49/Region: glycine-rich  
F:101-112/Region: alanine-rich  
F:162-180/Region: histidine/proline-rich  
F:186-201/Region: alanine-rich  
F:236-247/Region: glycine-rich  
F:267-291/Region: histidine/proline-rich  
F:316-383/Domain: POU domain homology <POU>  
F:402-458/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 1; Length 495;  
Best Local Similarity 84.6%; Pred. No. 27;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 186 AAAAAAAAXAA 198

## RESULT 28

JC5076  
myc-associated zinc-finger protein - human  
N;Alternate names: MAZ protein  
C;Species: Homo sapiens (man)  
C;Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 05-Nov-1999  
C;Accession: JC5076  
R;Tsuboi, H.; Sakatsume, O.; Itakura, K.; Yokoyama, K.K.  
Biochem. Biophys. Res. Commun. 226, 801-809, 1996  
A;Title: Members of the MAZ family: A novel cDNA clone for MAZ from human pancreatic islet  
A;Reference number: JC5076; MUID:96428591; PMID:8831693  
A;Accession: JC5076  
A;Molecule type: mRNA  
A;Residues: 1-497 <TSU>  
A;Cross-references: UNIPARC:UPI0000163839; DDBJ:D85131; NID:gi752741; PIDN:BAAL2728.1; EMBL:U00001  
A;Experimental source: pancreatic islet  
C;Comment: This protein plays a role in the control of transcriptional initiation of genes and between the introns of the mouse gene for immunoglobulin M-D.  
C;Keywords: phosphoprotein; zinc finger  
F;146,204,480/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F;349/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 100.0%; Score 44; DB 2; Length 497;  
Best Local Similarity 84.6%; Pred. No. 27;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 122 AAAAAAAAXAA 134

## RESULT 29

A48233  
polymavirus enhancer-binding protein 2 alpha chain type 1 - mouse  
N;Alternate names: PEA2 alpha chain type 1; PEA2 alpha chain type 2; PEBP2 alpha chain type 1  
C;Species: Mus musculus (house mouse)  
C;Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 09-Jul-2004  
C;Accession: A48233; B48233  
R;Ogawa, E.; Maruyama, M.; Kagoshima, H.; Inuzuka, M.; Lu, J.; Satake, M.; Shigesada, K.  
Proc. Natl. Acad. Sci. U.S.A. 90, 6859-6863, 1993  
A;Title: PEBP2/PEA2 represents a family of transcription factors homologous to the product of the PEA2 gene  
A;Reference number: A48233; MUID:93342088; PMID:8341710  
A;Accession: A48233  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-513 <OGA>  
A;Cross-references: UNIPROT:Q08775; UNIPARC:UPI000002B2F8; GB:D14636; NID:g391766; PIDN:BAAL2728.1; EMBL:U00001  
A;Accession: B48233  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-304, 'L', '306 <OG2>  
A;Cross-references: UNIPARC:UPI000002B2F9; GB:D14637; NID:g391768; PIDN:BAAL2728.1; EMBL:U00001  
C;Genetics:  
C;Superfamily: transcription factor CBF alpha 2  
C;Keywords: alternative splicing; DNA binding; T-cell; transcription factor; transcription factor

Query Match 100.0%; Score 44; DB 2; Length 513;  
Best Local Similarity 84.6%; Pred. No. 27;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 64 AAAAAAAAXAA 76

## RESULT 30

WJFFEN  
homeotic protein engrailed - fruit fly (Drosophila melanogaster)  
N;Alternate names: specific body pattern development protein  
C;Species: Drosophila melanogaster

C;Date: 27-Nov-1985 #sequence\_revision 27-Nov-1985 #text\_change 09-Jul-2004  
C;Accession: A90862; A93354; A03321; A25682; S03667  
R;Poole, S.J.; Kauvar, L.M.; Drees, B.; Kornberg, T.  
Cell 40, 37-43, 1985

A;Title: The engrailed locus of Drosophila: structural analysis of an embryonic transcription unit  
A;Reference number: A90862; MUID:85099327; PMID:3917855  
A;Accession: A90862  
A;Molecule type: mRNA  
A;Residues: 1-552 <POO>  
A;Cross-references: UNIPROT:P02836; UNIPARC:UPI000012CA13; GB:M10017; NID:gi57363; PIDN:BAAL2728.1; EMBL:U00001  
R;Fjose, A.; McGinnis, W.J.; Gehring, W.J.  
Nature 313, 284-289, 1985

A;Title: Isolation of a homobox-containing gene from the engrailed region of Drosophila  
A;Reference number: A93354; MUID:90114393; PMID:2481829  
A;Accession: A93354  
A;Molecule type: DNA  
A;Residues: 447-485, 'E', 487-518, 'WH' <EJO>  
A;Cross-references: UNIPARC:UPI00001745BF; GB:X01765; GB:K03059; NID:g8084; PIDN:CAA2590  
A;Note: the translation of the nucleotide sequence from Fig. 5 is shown, not the translation from Fig. 6

R;Kassis, J.A.; Poole, S.J.; Wright, D.K.; O'Farrell, P.H.  
EMBO J. 5, 3583-3589, 1986  
A;Title: Sequence conservation in the protein coding and intron regions of the engrailed gene  
A;Reference number: A91059; MUID:87161768; PMID:2881781  
A;Contents: annotation; intron locations and sequences  
R;Gav, N.J.; Poole, S.J.; Kornberg, T.B.  
Nucleic Acids Res. 16, 6637-6647, 1988

A;Title: The Drosophila engrailed protein is phosphorylated by a serine-specific protein kinase  
A;Reference number: S03667; MUID:88289425; PMID:2899884  
A;Contents: annotation; potential phosphorylation sites; homeobox domain  
C;Comment: This protein specifies the body segmentation pattern.  
C;Genetics:  
A;Gene: en

A;Cross-references: FlyBase:FBgn0000577  
A;Map position: 2R, 62.0 (48A1-4)  
A;Introns: 438/1; 470/3  
C;Superfamily: engrailed homeotic protein; homeobox; nucleus; segmentation; transcription regulation  
C;Keywords: DNA binding; embryo; homeobox; nucleus; segmentation; transcription regulation  
F;26-53/Region: glutamine-rich  
F;68-87/Region: alanine-rich  
F;232-240/Region: alanine-rich  
F;320-368/Region: serine-rich  
F;455-511/Domain: homeobox homology <HOK>

Query Match 100.0%; Score 44; DB 1; Length 552;  
Best Local Similarity 84.6%; Pred. No. 29;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
Db 68 AAAAAAAAXAA 80

## RESULT 31

A93369  
homeotic protein BarH1 - fruit fly (Drosophila ananassae)  
C;Species: Drosophila ananassae  
C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 31-Dec-2004  
C;Accession: A93369  
R;Kojima, T.; Ishimaru, S.; Higashijima, S.; Takayama, E.; Akimaru, H.; Sone, M.; Emori, Y.  
Proc. Natl. Acad. Sci. U.S.A. 88, 4343-4347, 1991

A;Title: Identification of a different-type homeobox gene, BarH1, possibly causing BarH1 phenotype  
A;Reference number: A93369; MUID:91239564; PMID:1674606  
A;Accession: A93369  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-604 <KOJ>  
A;Cross-references: UNIPROT:Q23838; UNIPARC:UPI0000075A37; GB:M59962; GB:M59963; NID:gi57363; PIDN:BAAL2728.1; EMBL:U00001  
C;Genetics:  
A;Gene: FlyBase:Dana/B

A;Cross-references: FlyBase:FBgn0012114  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;331-387/Domain: homeobox homology <HOK>

Query Match 100.0%; Score 44; DB 1; Length 552;  
Best Local Similarity 84.6%; Pred. No. 29;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Query Match 100.0%; Score 44; DB 2; Length 604;  
Best Local Similarity 84.6%; Pred. No. 31;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
DB 234 AAAAAAAASAA 246

RESULT 32  
S13367  
Om(1D) protein - fruit fly (Drosophila ananassae)  
C:Species: Drosophila ananassae  
C:Date: 18-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 05-Oct-2004  
C:Accession: S13367  
R:Tanda, S.; Corces, V.G.  
EMBO J. 10, 407-417, 1991  
A:Title: Retrotransposon-induced overexpression of a homeobox gene causes defects in eye  
A:Reference number: S13367; MUID:91122048; PMID:1671353  
A:Accession: S13367  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-606 <TAN>  
A:Cross-references: UNIPROT:P22544; UNIPARC:UPI000012C97A; EMBL:X56682; NID:g7146; PIDN:  
C:Genetics:  
A:Gene: FlyBase:Dana/B  
A:Cross-references: FlyBase:FBgn0012114  
A:Introns: 307/1; 383/2  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:332-388/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 44; DB 2; Length 606;  
Best Local Similarity 84.6%; Pred. No. 31;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
DB 234 AAAAAAAASAA 246

RESULT 33  
T02610  
probable YME1 ATP-dependant proteinase [imported] - Arabidopsis thaliana  
N:Alternate names: YTA11 protein homolog T19L18.5  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 31-Dec-2004  
C:Accession: T02610; H84656  
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC T19L18 genomic sequence.  
A:Reference number: Z14681  
A:Accession: T02610  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-627 <ROU>  
A:Cross-references: UNIPROT:O80983; UNIPARC:UPI000017885F; EMBL:AC004747; NID:g3413696;  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: H84656  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-627 <STO>  
A:Cross-references: UNIPARC:UPI000017885F; GB:AE002093; NID:g3413700; PIDN:AAC31223.1; G  
C:Genetics:  
A:Gene: T19L18.5; At2g26140  
A:Map position: 2  
A:Introns: 40/1; 119/1; 160/3; 319/3

C:Superfamily: Ftsh/SEC18/CDC48-type ATP-binding domain homology

Query Match 100.0%; Score 44; DB 2; Length 627;  
Best Local Similarity 84.6%; Pred. No. 32;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
DB 598 AAAAAAAASAA 610

RESULT 34  
S39356  
transcription factor btd - fruit fly (Drosophila sp.)  
C:Species: Drosophila sp.  
C:Date: 18-Feb-1994 #sequence\_revision 01-Dec-1995 #text\_change 07-May-1999  
C:Accession: S39356  
R:Wimmer, E.A.; Jaecckle, H.; Pfeifle, C.; Cohen, S.M.  
Nature 366, 690-694, 1993  
A:Title: A Drosophila homologue of human Spl is a head-specific segmentation gene.  
A:Reference number: S39356; MUID:94081952; PMID:8259212  
A:Accession: S39356  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <WIM>  
A:Cross-references: UNIPARC:UPI0000124C17; EMBL:Z29361; NID:g441283; PID:g441284  
C:Genetics:  
A:Gene: FlyBase:btd  
A:Cross-references: FlyBase:FBgn0000233  
A:Introns: 245/2

Query Match 100.0%; Score 44; DB 2; Length 644;  
Best Local Similarity 84.6%; Pred. No. 32;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
DB 205 AAAAAAAASAA 217

RESULT 35  
S40382  
box A-binding factor - fruit fly (Drosophila melanogaster)  
N:Alternate names: ABF; transcription factor dGATAB  
C:Species: Drosophila melanogaster  
C:Date: 13-Jan-1995 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C:Accession: S40382  
R:Abel, T.; Michelson, A.M.; Maniatis, T.  
Development 119, 623-633, 1993  
A:Title: A Drosophila GATA family member that binds to Adh regulatory sequences is expr  
A:Reference number: S40382; MUID:94244465; PMID:8187633  
A:Accession: S40382  
A:Molecule type: mRNA  
A:Residues: 1-779 <ABE>  
A:Cross-references: UNIPROT:P52172; UNIPARC:UPI0000135F5F; EMBL:X76217; NID:g441491; PID:  
C:Comment: This transcriptional activator is the earliest known marker of the developin  
C:Genetics:  
A:Gene: FlyBase:srp  
A:Cross-references: FlyBase:FBgn0003507  
C:Superfamily: box A-binding factor; GATA-type zinc finger homology  
C:Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zinc  
F:316-369/Domain: GATA-type zinc finger homology <GZF>  
F:319-343/Region: zinc finger GATA motif

Query Match 100.0%; Score 44; DB 1; Length 779;  
Best Local Similarity 84.6%; Pred. No. 37;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||  
DB 602 AAAAAAAASAA 614

## RESULT 36

JC7555  
C14orf4 protein - human  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004  
C:Accession: JC7555  
R:Rampazzo, A.; Fivotto, F.; Occhi, G.; Tiso, N.; Bortoluzzi, S.; Rowen, L.; Hood, L.; N  
Biochem. Biophys. Res. Commun. 278, 766-774, 2000  
A:Title: Characterization of C14orf4, a novel intronless human gene containing a polyglu  
A:Reference number: JC7555; MUID: 20549026; PMID:11095982  
A:Accession: JC7555  
A:Molecule type: DNA  
A:Residues: 1-796 <RAM>  
A:CROSS-references: UNIPROT:Q9H1B7; UNIPARC:UPI00000738BA; GB:AJ2777365  
C:Comment: This protein, a proline-rich nuclear protein, functions as a transcriptional  
s and undergoes rapid degradation.  
C:Genetics:  
A:Gene: C14orf4  
A:Map position: 14q24.3

Query Match 100.0%; Score 44; DB 2; Length 796;  
Best Local Similarity 84.6%; Pred. No. 38;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 88 AAAAAAAAXAA 100

## RESULT 37

T49385  
hypothetical protein B1D1.390 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: T49385  
R:Schulte, U.; Align, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: 225022  
A:Accession: T49385  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-805 <SCH>  
A:CROSS-references: UNIPROT:Q9P682; UNIPARC:UPI000017B46C; EMBL:AL355927; GSPDB:GN00116;  
A:Experimental source: BAC clone B1D1; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B1D1.390  
A:Map position: 6  
A:Introns: 54/3; 212/3

Query Match 100.0%; Score 44; DB 2; Length 805;  
Best Local Similarity 84.6%; Pred. No. 38;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 314 AATAAAAAAXAA 326

## RESULT 38

Tl3690  
hypothetical protein EG0003.2 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: Tl3690  
R:Murphy, L.; Harris, D.; Barrell, B.  
A:Submitted to the EMBL Data Library, November 1998  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17699  
A:Accession: Tl3690  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-806 <MUR>  
A:CROSS-references: UNIPROT:Q96828; UNIPARC:UPI000007B3C4; EMBL:AL031863; NID:e1331652;

## C:Genetics:

A:CROSS-references: FlyBase:FBgn0025833  
A:Introns: 37/3; 448/3; 611/2; 690/3  
A:Note: EG:EG0003.2

Query Match 100.0%; Score 44; DB 2; Length 806;  
Best Local Similarity 84.6%; Pred. No. 38;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 717 AAAAAAAAXAA 729

## RESULT 39

C88402  
protein H05C05.1 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: C88402  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: C88402  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-828 <STO>  
A:CROSS-references: UNIPROT:Q9TXU2; UNIPARC:UPI000007CE2B; GB:chr\_III; PIDN:AAC68790.1;  
C:Genetics:  
A:Gene: H05C05.1  
A:Map position: 3

Query Match 100.0%; Score 44; DB 2; Length 828;  
Best Local Similarity 84.6%; Pred. No. 39;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 358 AAAAAAAAXAA 370

## RESULT 40

A34692  
ecdysone-induced protein E74A - fruit fly (Drosophila melanogaster)  
N:Alternate names: ets-related protein E74A  
C:Species: Drosophila melanogaster  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 15-Mar-2004  
C:Accession: A34692  
R:Burtis, K.C.; Thummel, C.S.; Jones, C.W.; Karim, F.D.; Hogness, D.S.  
Cell 61, 85-99, 1990  
A:Title: The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene  
A:Reference number: A90912; MUID:90199900; PMID:2107982  
A:Accession: A34692  
A:Molecule type: mRNA  
A:Residues: 1-829 <BUR>  
A:CROSS-references: UNIPARC:UPI0000077909; GB:M37082; NID:g157307; PID:g157308  
C:Genetics:  
A:Gene: E74  
A:CROSS-references: FlyBase:FBgn0000567  
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation  
F:735-815/Domain: ets DNA-binding domain homology <ETS>

Query Match 100.0%; Score 44; DB 2; Length 829;  
Best Local Similarity 84.6%; Pred. No. 39;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 458 AAAAAAAAXAA 470

Search completed: September 9, 2006, 23:01:51  
Job time : 15.8101 secs

---

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:38:21 ; Search time 119.139 Seconds  
(without alignments)  
100.934 Million cell updates/sec

Title: US-10-617-568-36

Perfect score: 44

Sequence: 1 AAXAAAAAXAA 13

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

UniProt 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	100.0	52	Q5GZ88	ARATH
2	44	100.0	71	Q8NI53	homo sapien
3	44	100.0	85	ANP4_PSEAM	
4	44	100.0	91	ANPX_PSEAM	
5	44	100.0	112	Q3UUS5	MOUSE
6	44	100.0	115	Q3V062	ORYZA SATIV
7	44	100.0	117	Q3V062	MOUSE
8	44	100.0	130	Q9TUC9	CANFA
9	44	100.0	131	Q9TS88	FELCA
10	44	100.0	132	Q4TBK8	TEING
11	44	100.0	143	Q52HB4	CHICK
12	44	100.0	145	Q2XYX3	ANNPU
13	44	100.0	149	Q7Q8K9	ANOAGA
14	44	100.0	150	Q9Y4M1	HUMAN
15	44	100.0	153	Q765Q0	RAT
16	44	100.0	157	Q62GX4	BURMA
17	44	100.0	158	Q9MBF7	LILLO
18	44	100.0	164	Q7Q9W3	ANOAGA
19	44	100.0	164	Q8BNH5	MOUSE
20	44	100.0	167	Q765P9	RAT
21	44	100.0	171	Q8H8G6	ORYSA
22	44	100.0	175	Q5N9W5	ORYSA
23	44	100.0	191	Q96N13	HUMAN
24	44	100.0	199	Q4RK77	TEING
25	44	100.0	203	Q5SC14	HUMAN
26	44	100.0	213	Q3TUC8	MONDO
27	44	100.0	218	Q53G20	HUMAN
28	44	100.0	220	Q61PH7	HUMAN
29	44	100.0	223	Q5ISR0	MACFA
30	44	100.0	227	Q7Q706	ANOAGA
31	44	100.0	228	Q5K412	9MYRI

## ALIGNMENTS

32	44	100.0	228	2	Q5BQD5	9VIRU
33	44	100.0	228	2	Q5BQD8	9VIRU
34	44	100.0	228	2	Q5G7G9	9VIRU
35	44	100.0	233	1	ASCL11	RAT
36	44	100.0	233	1	ASCL11	ORYSA
37	44	100.0	236	1	ASCL11	HUMAN
38	44	100.0	236	2	Q6QDA1	HUMAN
39	44	100.0	236	2	Q9BX46	HUMAN
40	44	100.0	244	2	Q6CNB7	KJULA
41	44	100.0	247	2	Q9DDF6	PETMA
42	44	100.0	253	2	Q8WZ46	HUMAN
43	44	100.0	256	2	Q6L8B5	9DIPT
44	44	100.0	259	2	Q6L8B9	9DIPT
45	44	100.0	260	2	Q6L8B2	9DIPT
46	44	100.0	260	2	Q6L8B7	9DIPT
47	44	100.0	260	2	Q9ST44	DROME
48	44	100.0	266	2	Q6L8B8	9DIPT
49	44	100.0	266	2	Q9G3L0	DROME
50	44	100.0	267	2	Q6L173	DROME
51	44	100.0	268	2	Q6L8B6	9DIPT
52	44	100.0	269	2	Q4JHB6	HUMAN
53	44	100.0	272	2	Q753B0	ASHGO
54	44	100.0	276	1	SOX211	HUMAN
55	44	100.0	276	1	SOX211	MOUSE
56	44	100.0	276	2	Q5TBS1	HUMAN
57	44	100.0	277	2	Q68KY7	9DIPT
58	44	100.0	278	2	Q68KY5	9DIPT
59	44	100.0	279	2	Q68KY8	9DIPT
60	44	100.0	279	2	Q68KZ8	9DIPT
61	44	100.0	279	2	Q68L24	9DIPT
62	44	100.0	279	2	Q68L45	9DIPT
63	44	100.0	279	2	Q68L70	9DIPT
64	44	100.0	279	2	Q68L83	9DIPT
65	44	100.0	279	2	Q68L99	9DIPT
66	44	100.0	280	1	SOX211	CHICK
67	44	100.0	280	2	Q68L21	9DIPT
68	44	100.0	280	2	Q68L55	9DIPT
69	44	100.0	280	2	Q68L85	9DIPT
70	44	100.0	280	2	Q68L89	9DIPT
71	44	100.0	280	2	Q68LA1	9DIPT
72	44	100.0	281	2	Q68KZ6	9DIPT
73	44	100.0	281	2	Q68L03	9DIPT
74	44	100.0	281	2	Q68L43	9DIPT
75	44	100.0	281	2	Q68L53	9DIPT
76	44	100.0	281	2	Q68L79	9DIPT
77	44	100.0	281	2	Q68L81	9DIPT
78	44	100.0	282	2	Q68L04	9DIPT
79	44	100.0	282	2	Q68L07	9DIPT
80	44	100.0	282	2	Q68L12	9DIPT
81	44	100.0	282	2	Q68L15	9DIPT
82	44	100.0	282	2	Q68L16	9DIPT
83	44	100.0	282	2	Q68L73	9DIPT
84	44	100.0	282	2	Q68L84	9DIPT
85	44	100.0	282	2	Q68L90	9DIPT
86	44	100.0	282	2	Q68L91	9DIPT
87	44	100.0	282	2	Q68L94	9DIPT
88	44	100.0	282	2	Q68L96	9DIPT
89	44	100.0	282	2	Q68LA4	9DIPT
90	44	100.0	283	2	Q68KY6	9DIPT
91	44	100.0	283	2	Q68KZ0	9DIPT
92	44	100.0	283	2	Q68KZ4	9DIPT
93	44	100.0	283	2	Q68KZ5	9DIPT
94	44	100.0	283	2	Q68L02	9DIPT
95	44	100.0	283	2	Q68L18	9DIPT
96	44	100.0	283	2	Q68L19	9DIPT
97	44	100.0	283	2	Q68L44	9DIPT
98	44	100.0	283	2	Q68L47	9DIPT
99	44	100.0	283	2	Q68L50	9DIPT
100	44	100.0	283	2	Q68L58	9DIPT

Q5BQD5	mint virus
Q5BQD8	mint virus
Q5G7G9	mint virus
P19359	rattus norv
Q7XTV6	oryza sativ
P50553	homo sapien
Q6QDA4	homo sapien
Q9BX46	homo sapien
Q6CNB7	kluyveromyc
Q9DDF6	petromyzon
Q8WZ46	homo sapien
Q6L8B5	drosophila
Q6L8B9	drosophila
Q6L8B2	drosophila
Q9ST44	drosophila
Q6L8B8	drosophila
Q9G3L0	drosophila
Q6L173	drosophila
Q6L8B6	drosophila
Q4JHB6	homo sapien
Q753B0	ashbya goss
Q9Y651	homo sapien
Q8L1W0	mus musculus
Q5TBS1	homo sapien
Q68KY7	drosophila
Q68KY5	drosophila
Q68KY8	drosophila
Q68KZ8	drosophila
Q68L24	drosophila
Q68L45	drosophila
Q68L70	drosophila
Q68L83	drosophila
Q68L99	drosophila
Q9W7F5	gallus gall
Q68L21	drosophila
Q68L55	drosophila
Q68L85	drosophila
Q68L89	drosophila
Q68LA1	drosophila
Q68KZ6	drosophila
Q68L03	drosophila
Q68L43	drosophila
Q68L53	drosophila
Q68L79	drosophila
Q68L81	drosophila
Q68L04	drosophila
Q68L07	drosophila
Q68L12	drosophila
Q68L15	drosophila
Q68L16	drosophila
Q68L73	drosophila
Q68L84	drosophila
Q68L90	drosophila
Q68L91	drosophila
Q68L94	drosophila
Q68L96	drosophila
Q68LA4	drosophila
Q68KY6	drosophila
Q68KZ0	drosophila
Q68KZ4	drosophila
Q68KZ5	drosophila
Q68L02	drosophila
Q68L18	drosophila
Q68L19	drosophila
Q68L44	drosophila
Q68L47	drosophila
Q68L50	drosophila
Q68L58	drosophila



```

RESULT 1
Q56ZWS ARATH PRELIMINARY; PRT; 52 AA.
ID Q56ZWS8;
AC Q56ZWS8;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE FtsH like protease (Fragment).
GN Name=At2g26140;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Totoki Y., Seki M., Ishida J., Nakajima M., Enju A., Kamiya A.,
RA Narusaka M., Shin-i T., Nakagawa M., Sakamoto N., Oishi K., Kohara Y.,
RA Kobayashi M., Toyoda A., Sakaki Y., Sakurai T., Iida K., Akiyama K.,
RA Satou M., Toyoda T., Konagaya A., Carninci P., Kawai J.,
RA Hayaishizaki Y., Shinozaki K.
RA "Large-scale analysis of RIKEN Arabidopsis full-length (RAPL) cDNAs." ;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK220842; BAD94171.1; -; mRNA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001859; Ribosomal_P2.
DR PRINTS; PR00456; RIBOSOMALP2.
DR Protease.
KW Protease.
FT NON TER
SQ SEQUENCE 52 AA; 4750 MW; D8B358D27E6813B1 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 52;
Best Local Similarity 84.6%; Pred. No. 40;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 23 AAAAAAAAXAA 35

RESULT 2
Q8N153 HUMAN PRELIMINARY; PRT; 71 AA.
ID Q8N153;
AC Q8N153;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE RNA-binding protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Wang A., Gonzalez-Santos J.M., Hu J.;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF261889; AAM21972.1; -; Genomic DNA.
DR Ensembl; ENSG00000163694; Homo sapiens.
FT NON TER
SQ SEQUENCE 71 AA; 6958 MW; 3A0B336CD1048D99 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 71;
Best Local Similarity 84.6%; Pred. No. 52;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 45 AAAAAAAAXAA 57

RESULT 3
ANP4_PSEAM STANDARD; PRT; 85 AA.
ID ANP4_PSEAM
AC P02734;
DT 21-JUL-1986, integrated into UniProtKB/Swiss-Prot.
DT 13-AUG-1987, sequence version 1.
DT 07-FEB-2006, entry version 32.
DE Ice-structuring protein 4 precursor (ISP 4) (Antifreeze peptide 4).
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=81247379; PubMed=6265915;
RA Lin Y., Gross J.K.;
RT "Molecular cloning and characterization of winter flounder antifreeze
RT cDNA." ;
RL Proc. Natl. Acad. Sci. U.S.A. 78:2825-2829(1981).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point.
CC -!- SIMILARITY: Belongs to the type-I AFP family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; J00930; AAA49467.1; -; mRNA.
DR PIR; A03193; FDFL4W.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR Antifreeze protein; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 85
FT Ice-structuring protein 4.
FT /FTID=PRO 0000001684.
FT SEQUENCE 85 AA; 7215 MW; 8E62E1D2B4417BC CRC64;

Query Match 100.0%; Score 44; DB 1; Length 85;
Best Local Similarity 84.6%; Pred. No. 59;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 35 AATAAAAAAATAA 47

RESULT 4
ANPX_PSEAM STANDARD; PRT; 91 AA.
ID ANPX_PSEAM
AC P07835;
DT 01-AUG-1988, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1995, sequence version 2.
DT 07-FEB-2006, entry version 35.
DE Ice-structuring protein 2A7 precursor (ISP 2A7) (Antifreeze protein IIA7) (AFP).
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OX NCBI_TaxID=8265;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RX MEDLINE=81247379; PubMed=6265915;
RA Lin Y., Gross J.K.;
RT "Molecular cloning and characterization of winter flounder antifreeze
RT cDNA." ;
RL Proc. Natl. Acad. Sci. U.S.A. 78:2825-2829(1981).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point.
CC -!- SIMILARITY: Belongs to the type-I AFP family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; J00930; AAA49467.1; -; mRNA.
DR PIR; A03193; FDFL4W.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR Antifreeze protein; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 85
FT Ice-structuring protein 4.
FT /FTID=PRO 0000001684.
FT SEQUENCE 85 AA; 7215 MW; 8E62E1D2B4417BC CRC64;

Query Match 100.0%; Score 44; DB 1; Length 85;
Best Local Similarity 84.6%; Pred. No. 59;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 35 AATAAAAAAATAA 47
```

OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.  
 OX NCBI\_TaxID=8265;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [MRNA].  
 RX MEDLINE=85054993; PubMed=6548752;  
 RA Gourlie B., Lin Y., Price J., Devries A.L., Powers D., Huang R.C.C.;  
 RT "Winter flounder antifreeze proteins: a multigene family.";  
 RL J. Biol. Chem. 259:14960-14965(1984).  
 CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point.  
 CC -!- SIMILARITY: Belongs to the type-1 AFP family.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; M10148; AA049465.1; -; mRNA.  
 DR PIR; A22592; A22592.  
 DR InterPro; IPR000104; Antifreeze\_1.  
 DR PRINTS; PR00308; ANTIFREEZE1.  
 KW Antifreeze protein; Repeat; Signal.  
 KM  
 FT SIGNAL 1 21 Potential.  
 FT PROPEP 22 39 Removed by a dipeptidylpeptidase  
 FT (Probable).  
 FT /FTId=PRO\_0000001687.  
 FT Ice-structuring protein 2A7.  
 FT /FTId=PRO\_0000001688.  
 FT /FTId=PRO\_0000001689.  
 FT /FTId=PRO\_0000001690.  
 SQ SEQUENCE 91 AA; 8326 MW; DIFC542FD865012C CRC64;  
  
 Query Match 100.0%; Score 44; DB 1; Length 91;  
 Best Local Similarity 84.6%; Pred. No. 63;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
 Qy 1 AAXAAAAAAAXAA 13  
 ||| ||||| ||  
 Db 52 AATAAAAAAATAA 64  
  
 RESULT 5  
 Q3UUS5\_MOUSE  
 ID Q3UUS5\_MOUSE PRELIMINARY; PRT; 112 AA.  
 AC Q3UUS5;  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 07-FEB-2006, sequence version 1.  
 DE 16 days neonate thymus cDNA, RIKEN full-length enriched library.  
 DE clone:AL30089M21 product:HBV px associated protein 8 large isoform  
 DE homolog.  
 GN Name=Rbxap;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX STRAIN=C57BL/6J; TISSUE=Thymus;  
 RC MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX STRAIN=C57BL/6J; TISSUE=Thymus;  
 RC PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzi R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturalien R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dallinga B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA

RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelson J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Motagani-Ribar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reid J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Bruscia V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Nimomiya N.,  
 RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Oshino T., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tonita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX

RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RA "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tegami M., Tagami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RA Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AK138087; BAE23549.1; -; mRNA.  
 DR MGI; MGI:2682305; Hbxap.  
 SQ SEQUENCE 112 AA; 11657 MW; 3F364EA14DDFD076 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 112;  
 Best Local Similarity 84.6%; Pred. No. 74;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Q/ 1 AAXAAAAAAAXAA 13

Db 4 AAAAAAAAXAA 16

RESULT 6

Q656T8\_ORYSA

ID Q656T8\_ORYSA PRELIMINARY; PRT; 115 AA.  
 AC Q656T8;  
 DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.  
 DT 25-OCT-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 8.  
 DE Cell division protein (FtsH)-like.  
 DE Name=B1151A10.24-1;  
 GN Oryza sativa (japonica cultivar-group).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BPF clade;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22337376; PubMed=12447438; DOI=10.1038/nature01184;  
 RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,  
 RA Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
 RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
 RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
 RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
 RA Ikono M., Ito S., Ito T., Ito Y., Iwabuchi A., Kamiya K.,  
 RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,  
 RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,  
 RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
 RA Naniki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,  
 RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,  
 RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshinara R., Yukawa K.,  
 RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.-I., Eun M.-Y.,  
 RA Yano M., Jiang J., Gojobori T.;  
 RT "The genome sequence and structure of rice chromosome 1.";  
 RL Nature 420:312-316(2002).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AP003413; BAD45189.1; -; Genomic DNA.  
 DR Gramene; Q656T8;  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004222; P:metalloendopeptidase activity; IEA.  
 DR GO; GO:0001301; P:cell division; IEA.  
 DR GO; GO:0006508; P:proteolysis; IEA.  
 DR InterPro; IPR000642; Peptidase\_M41.  
 DR Pfam; PF01434; Peptidase\_M41; 1.  
 KW Cell division.  
 SQ SEQUENCE 115 AA; 11813 MW; B64A595C9756DEC5 CRC64;  
 Query Match 100.0%; Score 44; DB 2; Length 115;  
 Best Local Similarity 84.6%; Pred. No. 75;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Q/ 1 AAXAAAAAAAXAA 13  
 Db 85 AAAAAAAAXAA 97  
 RESULT 7  
 Q3V062\_MOUSE PRELIMINARY; PRT; 117 AA.  
 AC Q3V062;  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 11-OCT-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Adult male testis cDNA, RIKEN full-length enriched library,  
 DE clone:4933416E08 product:hypothetical Proline-rich region  
 DE profile/Alanine-rich region profile containing protein, full insert  
 DE sequence.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX PubMed=16141073; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
Baisel K., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
Hill D., Humnietek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
Kurochkin I.V., Larau L.F., Lazarevic D., Lipovich L., Liu J.,  
Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
Mottaqui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G.,  
Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
Roat B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
Schonbach C., Sekiguchi K., Sempke C.A., Seno S., Sessa L.,  
Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
Sperling S., Stupka E., Sugiuira K., Sultana R., Takenaka Y., Taki K.,  
Tamao J., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
Yamanashi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
(Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
Yagi K., Tomaru Y., Haegawa Y., Nogami A., Schonbach C., Gojobori T.,  
Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,

Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
Sultana R., Tanaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang L., Yang L.,  
Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
Hirokawa T., Kawai J., Kawai J., Kawai K., Arakawa T., Fukuda S.,  
Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,  
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-cropper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RX MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Testis;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
Shibata K., Shiraki T., Tagami M., Tagami M., Waki K., Watahiki A.,  
Muramatsu M., Hayashizaki Y.;  
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RL

```

CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK133410; BAB21643.1; -; mRNA.
KW Hypothetical protein.
SQ SEQUENCE 117 AA; 12609 MW; 61282448B68F9365 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 117;
Best Local Similarity 84.6%; Pred. No. 76;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 22 AASAAAAAAAXAA 34

RESULT 8
Q9TUC9_CANFA PRELIMINARY; PRT; 130 AA.
AC Q9TUC9;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-MAY-2000, sequence version 1.
DT 07-FEB-2006, entry version 15.
DE Transcription factor HOXA13 (Fragment).
GN Name=HOXA13;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20122168; PubMed=10656931; DOI=10.1007/s003350010029;
RA Mortlock D.P., Sateesh P., Innis J.W.;
RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein.";
RL Mamm. Genome 11:151-158(2000).
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF083096; AAD54641.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001859; Ribosomal_P2.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Nuclear protein.
FT NON TER 1
FT NON TER 130
SQ SEQUENCE 130 AA; 10975 MW; CCCA1A7B52221394 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 130;
Best Local Similarity 84.6%; Pred. No. 83;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 16 AAAAAAAAXAA 28

RESULT 9
Q9TST8_FELCA PRELIMINARY; PRT; 131 AA.
ID Q9TST8_FELCA
AC Q9TST8;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-MAY-2000, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Transcription factor HOXA13 (Fragment).
GN Name=HOXA13;
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Euthera; Laurasiatheria; Carnivora; Fissipedia; Felidae;
OC Felinae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20122168; PubMed=10656931; DOI=10.1007/s003350010029;
RA Mortlock D.P., Sateesh P., Innis J.W.;
RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein.";
RL Mamm. Genome 11:151-158(2000).
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF083095; AAD54640.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001859; Ribosomal_P2.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Nuclear protein.
FT NON TER 1
FT NON TER 131
SQ SEQUENCE 131 AA; 11046 MW; 950562B80E529D4F CRC64;

Query Match 100.0%; Score 44; DB 2; Length 131;
Best Local Similarity 84.6%; Pred. No. 83;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
Db 16 AAAAAAAAXAA 28

RESULT 10
Q4TBK8_TETNG PRELIMINARY; PRT; 132 AA.
ID Q4TBK8_TETNG
AC Q4TBK8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Chromosome end determined SCAF7119, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00003716001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15496914; DOI=10.1038/nature03025;
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,
RA Maucell E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bionnet C., Skalli Z., Cattolico L., Poulain J., De Bernardis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,
RA Parra M., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an

```

```
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAB01007119; CAP99724.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 132
SQ SEQUENCE 132 AA; 12324 MW; 2761624E77ABA2CF CRC64;

Query Match 100.0%; Score 44; DB 2; Length 132;
Best Local Similarity 84.6%; Pred. No. 84;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13
Db 84 AAAAAAAAAAAAAA 96

RESULT 11
Q52HB4 CHICK
ID Q52HB4 CHICK PRELIMINARY; PRT; 143 AA.
AC Q52HB4;
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 24-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Pou3f3 (Fragment).
GN Name=Brnl;
OS Gallus Gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archoosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15872005; DOI=10.1242/dev.01833;
RA Bouchard M., Grote D., Craven S.E., Sun Q., Steinlein P.,
RA Busslinger M.;
RT "Identification of Pax2-regulated genes by expression profiling of the
RT mid-hindbrain organizer region.";
RL Development 132:2633-2643 (2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DQ002393; AAY17903.1; -; mRNA.
FT NON_TER 1
FT NON_TER 143
SQ SEQUENCE 143 AA; 14137 MW; 5C1818C8EC2121F4 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 143;
Best Local Similarity 84.6%; Pred. No. 89;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13
Db 33 AAAAAAAAAAAAAA 45

RESULT 12
Q2YXK3 ANNPV
ID Q2YXK3 ANNPV PRELIMINARY; PRT; 145 AA.
AC Q2YXK3;
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 20-DEC-2005, sequence version 1.
DT 21-FEB-2006, entry version 4.
DE HOXA13 (Fragment).
GN Name=HOXA13;
OS Anniella pulchra (California legless lizard).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scieroglossa; Anguimorpha; Anguillae;
OC Anniellinae; Anniella.
```

```
OX NCBI_TaxID=51957;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=16292255;
RA Fry B.G., Vidal N., Norman J.A., Vonk F.J., Scheib H., Ramjan S.F.,
RA Kuruppu S., Fung K., Blair Hedges S., Richardson M.K., Hodgson W.C.,
RA Ignjatovic V., Summerhayes R., Kochva E.;
RT "Early evolution of the venom system in lizards and snakes.";
RL Nature 439:584-588 (2006).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DQ119605; ABA41582.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 145
SQ SEQUENCE 145 AA; 14182 MW; A2AEAF2330637142 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 145;
Best Local Similarity 84.6%; Pred. No. 90;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13
Db 14 AAAAAAAAAAAAAA 26

RESULT 13
Q7Q8K9 ANOGA
ID Q7Q8K9 ANOGA PRELIMINARY; PRT; 149 AA.
AC Q7Q8K9;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 2.
DT 07-FEB-2006, entry version 13.
DE ENSANGP0000020819.
GN ORFNames=ENSANGG00000018330;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAB01008944; EAA10278.2; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001699; TF_T-box.
DR PANTHER; PTHR11267; TF_T-box; 1.
DR Pfam; PF00907; T-box; 1.
DR PROSITE; PS0252; TBOX 3; 1.
SQ SEQUENCE 149 AA; 15582 MW; 6EA2E604F4ED59C CRC64;

Query Match 100.0%; Score 44; DB 2; Length 149;
Best Local Similarity 84.6%; Pred. No. 92;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 AAXAAAAAAXAA 13
Db 10 AAAAAAAXAAAA 22

RESULT 14
Q9Y4M1_HUMAN PRELIMINARY; PRT; 150 AA.
AC Q9Y4M1
DT 01-NOV-1999, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1999, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE Hypothetical protein DKF2P586E1621 (Fragment).
GN Name=DKF2P586E1621;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Uterus;
RC The German cDNA Consortium;
RG Ottenwaelder B.; Obermaier B.; Deutschenbaur S.; Schaipp A.;
RA Mewes H.W.; Weil B.; Anid C.; Osanger A.; Fobo G.; Han M.; Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL080235; CAB45781.1; -; mRNA.
DR PIR; T12547; T12547.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 150 AA; 15110 MW; B0C80E466FCAB03E CRC64;

Query Match 100.0%; Score 44; DB 2; Length 150;
Best Local Similarity 84.6%; Pred. No. 93;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAXAA 13
Db 129 AAAAAAAXAAAA 141

RESULT 15
Q765Q0_RAT PRELIMINARY; PRT; 153 AA.
AC Q765Q0
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE Core binding factor alpha subunit type I (Fragment).
GN Name=CBFAl;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP TISSUE=Osteoblast;
RC Dohi Y.; Ohgushi H.; Nakajima H.; Ikeuchi M.; Shimaoka H.;
RA Yonemasa K.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB115745; BAD08305.1; -; mRNA.
DR SMR; Q765Q0; 93-153.
DR Ensembl; ENSRNOG0000020193; Rattus norvegicus.
DR GO; GO:0005634; C:nucleus; IEA.
```

```
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000040; AML1_Runt.
DR InterPro; IPR012346; PS3_RUNT_DNA_bd.
DR PANTHER; PTHR11950; AML1_Runt; 1.
DR Pfam; PF00853; Runt; 1.
DR PRINTS; PR00967; ONCOGENEAML1.
DR PROSITE; PS51062; RUNT; 1.
FT NON_TER 153
SQ SEQUENCE 153 AA; 16521 MW; D8A2B1598C1EC364 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 153;
Best Local Similarity 84.6%; Pred. No. 94;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAXAA 13
Db 65 AAAAAAAXAAAA 77

RESULT 16
Q62GX4_BURMA PRELIMINARY; PRT; 157 AA.
AC Q62GX4
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Acetyl-CoA carboxylase, biotin carboxyl carrier protein.
GN Name=accB; OrderedLocusNames=BMA2501;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 23344; DOI=10.1073/pnas.0403306101;
RX PubMed=15377793;
RA Niemman W.C.; Deshaizer D.; Kim H.S.; Tettelin H.; Nelson K.E.;
RA Feldblyum T.V.; Ulrich R.L.; Ronning C.M.; Brinkac L.M.;
RA Daugherty S.C.; Davidson T.D.; DeBoy R.T.; Dimitrov G.; Dodson R.J.;
RA Durkin A.S.; Gwinn M.L.; Haft D.H.; Khouri H.M.; Kolonay J.F.;
RA Madupu R.; Mohammed Y.; Nelson W.C.; Radune D.; Romero C.M.;
RA Sarria S.; Selengut J.; Shambhlin C.; Sullivan S.A.; White O.; Yu Y.;
RA Zafar N.; Zhou L.; Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000010; AAU50135.1; -; Genomic_DNA.
DR TIGR; BMA2501;
DR GO; GO:0009317; C:acetyl-CoA carboxylase complex; IEA.
DR GO; GO:0003989; F:acetyl-CoA carboxylase activity; IEA.
DR GO; GO:0009374; F:biotin binding; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR InterPro; IPR001249; AcCoA_biotinCC.
DR InterPro; IPR001882; Biotin_BS.
DR InterPro; IPR000089; Biotin_lipoyl.
DR Pfam; PF00364; Biotin_lipoyl; 1.
DR PRINTS; PR01071; AC0ABIOTINCC.
DR TIGRFAMs; TIGR00531; BCCP; 1.
DR PROSITE; PS00188; BIOTIN; 1.
DR PROSITE; PS00968; BIOTINYL_LIPOYL; 1.
DR Biotin; Complete proteome.
SQ SEQUENCE 157 AA; 16082 MW; A0B41233C5AE51B0 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 157;
Best Local Similarity 84.6%; Pred. No. 96;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAXAA 13
```

```
Db      64 AAAAAAAAAAAAAA 76
|| ||||| ||
|| ||||| ||

RESULT 17
Q9MBF7 LILLO PRELIMINARY; PRT; 158 AA.
AC Q9MBF7 LILLO
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2000, sequence version 1.
DT 21-FEB-2006, entry version 25.
DE GH2B.
OS Lilium longiflorum (Trumpet lily).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Liliaceae; Lilium.
OX NCBI_TaxID=4690;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=20252564; PubMed=10794571; DOI=10.1007/s004120050401;
RA Ueda K., Kinoshita Y., Xu Z.-J., Ide N., Ono M., Akahori Y.,
RT "Unusual core histones specifically expressed in male gametic cells of
RT Lilium longiflorum.";
RL Chromosome 108:491-500(2000).
CC -!- SUBUNIT: The nucleosome is an octamer containing two molecules
CC each of H2A, H2B, H3 and H4. The octamer wraps approximately 146
CC bp of DNA (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB003780; BAA96095.1; -; mRNA.
DR HSSP; P02281; 1KK3.
DR GO; GO:0005694; C:chromosome; IEA.
DR GO; GO:0000786; C:nucleosome; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0007001; P:chromosome organization and biogenesis (sen. .; IEA.
DR GO; GO:0006334; P:nucleosome assembly; IEA.
DR InterPro; IPR007124; Hist_TAF.
DR InterPro; IPR007125; Histone_core_D.
DR InterPro; IPR005558; Histone_H2B.
DR PANTHER; PTHR11425; Histone_H2B; 1.
DR PANTHER; PTHR11425:SF4; Histone_H2B; 1.
DR Pfam; PF00125; Histone; 1.
DR PRINTS; PR00621; HISTONEH2B.
DR ProDom; PD000497; Histone_H2B; 1.
DR SMART; SM00427; H2B; 1.
DR KW Chromosomal protein; DNA-binding; Nuclear protein; Nucleosome core.
SQ SEQUENCE 158 AA; 17249 MW; 37A96BD9117A0BD CRC64;

Query Match 100.0%; Score 44; DB 2; Length 158;
Best Local Similarity 84.6%; Pred. No. 96;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAXAAAAAAAAAXAA 13
|| ||||| ||
|| ||||| ||

RESULT 18
Q7Q9M3 ANOGA PRELIMINARY; PRT; 164 AA.
AC Q7Q9M3 ANOGA
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 2.
DT 07-FEB-2006, entry version 8.
DE ENSANGP00000015660 (fragment).
GN ANPNames=ENSANGG00000013171;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
```

```
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAB01008900; BAA09564.2; -; Genomic_DNA.
DR FT NON TER 164 164
SQ SEQUENCE 164 AA; 16803 MW; 2D1BAC62C087BBAC CRC64;

Query Match 100.0%; Score 44; DB 2; Length 164;
Best Local Similarity 84.6%; Pred. No. 99;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 AAXAAAAAAAAAXAA 13
|| ||||| ||
|| ||||| ||

RESULT 19
Q8BNH5 MOUSE PRELIMINARY; PRT; 164 AA.
AC Q8BNH5;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE 9 days embryo whole body cDNA, RIKEN full-length enriched library,
DE clone: D030067F24 product: hypothetical protein, full insert sequence.
OS Mus musculus (Mouse).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Euthera; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Whole body;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Whole body;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RX Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RX Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RX Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RX Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RX Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RX Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RX Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RX Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RX di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RX Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RX Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RX Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RX Hill D., Humincicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RX Jakt M., Kanapin A., Katoh M., Kawasaki Y., Kelso J., Kitamura H.,
RX Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
```



RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda K., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlandi V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.B., Sadelin A., Schneider C., Sheng Y.,  
RA Schonbach C., Sekiguchi K., Sempile C.A., Seno S., Sessa L.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrester A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kenai A., Kawai J., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sadelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staib F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schonbach C., Seya T., Shibata Y., Sorch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Yoneda Y., Ishikawa T., Ogawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Whole body;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AK083689; BAC38994.1; -; mRNA.  
DR Ensembl: ENSMUSG00000015882; Mus musculus.  
KW Hypothetical protein.  
SQ SEQUENCE 164 AA; 18462 MW; FD764AE704780A28 CRC64;  
  
Query Match 100.0%; Score 44; DB 2; Length 164;  
Best Local Similarity 84.6%; Pred. No. 99;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAXAA 13  
DB 9 AAAAAAAAAAAAA 21  
  
RESULT 20

Q765P9 RAT  
ID Q765P9 RAT PRELIMINARY; PRT; 167 AA.  
AC Q765P9;  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 9.  
DE Core binding factor alpha1 subunit type II (Fragment).  
GN Name=CBFA1;  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Osteoblast;  
RA Dohi Y., Ohgushi H., Nakajima H., Ikeuchi M., Shimaoka H.,  
RA Yonemasa K.;  
RL Submitted (JUL-2003) to the EMBL/GenBank/DBSJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AB115746; BAD08306.1; -; mRNA.  
DR SMR; Q765P9; 107-167.  
DR GO; ENSRNOG0000020193; Rattus norvegicus.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0003700; F:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR000040; AMUL\_Runt.  
DR InterPro; IPR012346; P53\_RUNT\_DNA\_bd.  
DR PANTHER; PTHR11950; AMUL\_Runt; 1.  
DR Pfam; PF00853; Runt; 1.  
DR PRINTS; PR00967; ONCOGENEAM1.  
DR PROSITE; PS51062; RUNT; 1.  
FT NON\_TER 167 167  
SQ SEQUENCE 167 AA; 18085 MW; E3C437F8F04BE74A CRC64;  
  
Query Match 100.0%; Score 44; DB 2; Length 167;  
Best Local Similarity 84.6%; Pred. No. 1e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 79 AAAAAAAAXAA 91  
  
RESULT 21  
Q8H8G6 ORYSA  
ID Q8H8G6 ORYSA PRELIMINARY; PRT; 171 AA.  
AC Q8H8G6;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 21-FEB-2006, entry version 10.  
DE Putative KNOX family class 2 homeodomain protein.  
GN Name=OJ1126B12.15;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
OC Ehrhartoideae; Oryzaeae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,  
RA Currie J., Collura K.;  
RT "Rice Genomic Sequence";  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AC098695; AAN74840.1; -; Genomic\_DNA.

DR Gramene; Q8H8G6; -;  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR InterPro; IPR005540; KNOX1.  
DR InterPro; IPR005541; KNOX2.  
DR Pfam; PF03790; KNOX1; 1.  
DR Pfam; PF03791; KNOX2; 1.  
KW Nuclear protein.  
SQ SEQUENCE 171 AA; 17982 MW; 647C830E1C22B8F5 CRC64;  
  
Query Match 100.0%; Score 44; DB 2; Length 171;  
Best Local Similarity 84.6%; Pred. No. 1e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 24 AAAAAAAAXAA 36  
  
RESULT 22  
Q5N9W5 ORYSA  
ID Q5N9W5 ORYSA PRELIMINARY; PRT; 175 AA.  
AC Q5N9W5;  
DT 01-FEB-2005, integrated into UniProtKB/TrEMBL.  
DT 01-FEB-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE Hypothetical protein P0406G08.29.  
GN Name=P0406G08.29;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
OC Ehrhartoideae; Oryzaeae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22337376; PubMed=12447438; DOI=10.1038/nature01184;  
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,  
RA Wu J., Nilimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
RA Ikono M., Ito S., Ito Y., Ito Y., Iwabuchi A., Kamiya K.,  
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,  
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,  
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,  
RA Shimokawa T., Shomura A., Song J., Takasaki Y., Terasawa K., Tsuji K.,  
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,  
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.-I., Eun M.-Y.,  
RA Yano M., Jiang J., Gojobori T.;  
RT "The genome sequence and structure of rice chromosome 1";  
RL Nature 420:312-316(2002).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AP003240; BAD81742.1; -; Genomic\_DNA.  
DR Gramene; Q5N9W5; -;  
KW Hypothetical protein.  
SQ SEQUENCE 175 AA; 18520 MW; 2FF4E34C0E89871 CRC64;  
  
Query Match 100.0%; Score 44; DB 2; Length 175;  
Best Local Similarity 84.6%; Pred. No. 1e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 10 AAAAAAAAXAA 22  
  
RESULT 23  
Q96NI3 HUMAN  
ID Q96NI3\_HUMAN PRELIMINARY; PRT; 191 AA.

AC Q96N13;  
DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
DT 01-DEC-2001, sequence version 1.  
DT 07-MAR-2006, entry version 21.  
DE Hypothetical protein FLJ30829 (Hypothetical protein RBM24)  
DE (OTTHUMP0000016067).  
GN Name=RBM24; ORFNames=RP1-259A10.1-002;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamiyama K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hirao K., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yoshida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shionata N., Sano S.,  
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RL Nat. Genet. 36:40-45 (2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallyah S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Huiyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kertanen M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Brain;  
RG NIH MGC Project;  
RL Submitted (SEP-2005) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RA Babbage A.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK055391; BAB70914.1; -; mRNA.  
DR EMBL; BC104808; AAI04809.1; -; mRNA.  
DR EMBL; BC104810; AAI04811.1; -; mRNA.  
DR EMBL; AL136305; CAI19958.1; -; Genomic DNA.  
DR Ensembl; ENSG00000112183; Homo sapiens.  
DR InterPro; IPR001778; POA\_allergen\_C.  
DR PRINTS; PR00813; POAALLERGEN.  
DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_I; UNKNOWN\_1.  
KW Hypothetical protein.  
SQ SEQUENCE 191 AA; 19599 MW; 8954D14E3DA941F7 CRC64;  
Query Match 100.0%; Score 44; DB 2; Length 191;  
Best Local Similarity 84.6%; Pred. No. 1.1e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAXAAAAAAAXAA 13  
Db 164 AAAAAAAAXAAAA 176  
RESULT 24  
Q4RK77 TETNG  
ID Q4RK77 TETNG PRELIMINARY; PRT; 199 AA.  
AC Q4RK77  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, sequence version 3.  
DE Chromosome 2, SCF15032, whole genome shotgun sequence. (Fragment).  
GN ORFNames=GSTENG0003100001;  
OS Tetraodon nigroviridis (Green puffer).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthopterygia; Acanthopterygii; Percormorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Tetraodon.  
OX NCBI\_TaxID=99883;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15496914; DOI=10.1038/nature03025;  
RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dessat C., Segurens B.,  
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
RA Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
RA Bionnet C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
RA Cruaud C., Duprat G., Chappie C., Coutancu J.-P., Guzy J.,  
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,  
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,  
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
RA Laudet V., Schachner V., Quetier F., Saurin W., Scarpelli C.,  
RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;  
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
RT the early vertebrate proto-karyotype";  
RN Nature 431:946-957 (2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RG Genoscope; Whitehead Institute Centre for Genome Research;  
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----

```
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAB01015032; CAG11205.1; -; Genomic_DNA.
FT NON_TER 1
SQ SEQUENCE 199 AA; 21690 MW; FB839A4F470FFD5D CRC64;

Query Match 100.0%; Score 44; DB 2; Length 199;
Best Local Similarity 84.6%; Pred. No. 1.2e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAXAAXAAXAA 13
||| ||||| |||
Db 147 AAXAAXAAXAAXAA 159

RESULT 25
Q5SC14 HUMAN PRELIMINARY; PRT; 203 AA.
AC Q5SC14
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Basic helix-loop-helix protein 3 isoform 2.
GN Name=BHLHB3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RA Falvella F.S., Manenti G., Spinola M., Pignatiello C., Ravagnani F.,
RA Conti B., Pastorino U., Dragani T.A.;
RT "Population-based association study on two candidate lung
RT adenocarcinoma modifier genes flanking the D12S1034 locus.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY665466; AAV54601.1; -; mRNA.
DR Ensembl; ENSG00000123095; Homo sapiens.
SQ SEQUENCE 203 AA; 19438 MW; 02625F0B557CB263 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 203;
Best Local Similarity 84.6%; Pred. No. 1.2e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAXAAXAAXAA 13
||| ||||| |||
Db 117 AAXAAXAAXAAXAA 129

RESULT 26
Q9TUC8 MONDO PRELIMINARY; PRT; 213 AA.
AC Q9TUC8;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-MAY-2000, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Transcription factor HOXA13 (Fragment).
GN Name=HOXA13;
OS Monodelphis domestica (Short-tailed gray opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
OX NCBI_TaxID=13616;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20122168; PubMed=10656931; DOI=10.1007/s003350010029;
RA Mortlock D.P., Sateesh P., Innis J.W.;
RT "Evolution of N-terminal sequences of the vertebrate HOXA13 protein.";
```

```
RL Mamm. Genome 11:151-158(2000).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF083097; AAD54642.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
KW Nuclear protein.
FT NON_TER 1
FT NON_TER 213
SQ SEQUENCE 213 AA; 19165 MW; EDEC8B40FCACADE9 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 213;
Best Local Similarity 84.6%; Pred. No. 1.2e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAXAAXAAXAA 13
||| ||||| |||
Db 16 AAXAAXAAXAAXAA 28

RESULT 27
Q53G20 HUMAN PRELIMINARY; PRT; 218 AA.
AC Q53G20;
DT 24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 24-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Ribosomal protein L14 variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94171032; PubMed=8125298; DOI=10.1016/0378-1119(94)90802-8;
RA Maruyama K., Sugano S.;
RT "Oligo-capping: a simple method to replace the cap structure of
RT eucaryotic mRNAs with oligoribonucleotides.";
RL Gene 138:171-174 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98038986; PubMed=9373149; DOI=10.1016/S0378-1119(97)00411-3;
RA Suzuki Y., Yoshitomo K., Maruyama K., Sugano A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
RT end-enriched cDNA library.";
RL Gene 200:149-156 (1997).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK223111; BAD96831.1; -; mRNA.
DR Ensembl; ENSG00000139239; Homo sapiens.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
KW Ribosomal protein.
FT NON_TER 1
SQ SEQUENCE 218 AA; 23645 MW; 3DE5CB1F97345E78 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 218;
Best Local Similarity 84.6%; Pred. No. 1.2e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAXAAXAAXAA 13
||| ||||| |||
```

```
Db 150 AAAAAAAAAAAAAA 162

RESULT 28
Q61PH7_HUMAN PRELIMINARY; PRT; 220 AA.
AC Q61PH7;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DE 07-MAR-2006, entry version 8.
DE RPL14 protein (Ribosomal protein L14 variant).
GN Name=RPL14;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RA Director MGC Project;
RA Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Dermoid cancer;
RX MEDLINE=94171032; PubMed=8125298; DOI=10.1016/0378-1119(94)90802-8;
RA Maruyama K., Sugano S.;
RT "Oligo-capping : a simple method to replace the cap structure of
eucaryotic mRNAs with oligoribonucleotides.";
RL Gene 138:171-174(1994).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Dermoid cancer;
RX MEDLINE=98038986; PubMed=9373149; DOI=10.1016/S0378-1119(97)00411-3;
RA Suzuki Y., Yoshitomo K., Maruyama K., Suyama A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
end-enriched cDNA library.";
RL Gene 200:149-156(1997).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Dermoid cancer;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC071913; AAT71913.1; -; mRNA.

DR EMBL; AK222750; BAD96470.1; -; mRNA.
DR Ensembl; ENSG00000139239; Homo sapiens.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR002784; Ribosomal_L14e.
DR Pfam; PF00467; KOW; 1.
DR Pfam; PF01929; Ribosomal_L14e; 1.
KW Ribosomal protein.
SQ SEQUENCE 220 AA; 23787 MW; 0FD8AA7FC191E864 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 220;
Best Local Similarity 84.6%; Pred. No. 1.2e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAXXAA 13
||| ||||| ||
Db 150 AAAAAAAAAAAAAA 162

RESULT 29
Q51SR0_MACFA PRELIMINARY; PRT; 223 AA.
AC Q51SR0;
DT 15-FEB-2005, integrated into UniProtKB/TrEMBL.
DT 15-FEB-2005, sequence version 1.
DE 07-FEB-2006, entry version 6.
DE Achaete-scute complex protein 1 (Fragment).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
OC Cercopithecidae; Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15820360; DOI=10.1016/j.cell.2004.11.040;
RA Dorus S., Vallender E.J., Evans P.D., Anderson J.R., Gilbert S.L.,
RA Mahowald M., Wyckoff G.J., Malcom C.M., Lahn B.F.;
RT "Accelerated evolution of nervous system genes in the origin of Homo
sapiens.";
RL Cell 119:1027-1040(2004).
CC -i- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AY650320; AAV67352.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0030528; F:transcription regulator activity; IEA.
DR GO; GO:0045449; P:regulation of transcription; IEA.
DR InterPro; IPR001092; HLH basic.
DR InterPro; IPR011598; HLH_DNA_bd.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS50888; HLH; 1.
FT NON_TER 1
FT NON_TER 223
FT NON_TER 223
SQ SEQUENCE 223 AA; 23804 MW; CD0D8E7C9F4F1063 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 223;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAXXAA 13
||| ||||| ||
Db 34 AAAAAAAAAAAAAA 46

RESULT 30
Q7Q706_ANOGA PRELIMINARY; PRT; 227 AA.
ID Q7Q706 ANOGA
AC Q7Q706;
```

15-DEC-2003, integrated into UniProtKB/TrEMBL.  
 07-DEC-2004, sequence version 2.  
 07-FEB-2006, entry version 10.  
 ENSANGP0000017596 (Fragment).  
 GN ORFNames=ENSANG0000015107;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;  
 OC Anophelinae; Anopheles.  
 OX NCBI\_TaxID=180454;  
 [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=PEST;  
 RC The Anopheles gambiae Sequence Committee;  
 RG "Anopheles gambiae re-annotation."  
 RT Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.  
 RL [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=PEST;  
 RC The Anopheles gambiae Sequence Committee;  
 RG Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL: AAB01008960; EAA11077.2; -; Genomic\_DNA.  
 DR GO: 0005634; C:nucleus; IEA.  
 DR GO: 0003676; F:nucleic acid binding; IEA.  
 DR GO: 0008270; F:zinc ion binding; IEA.  
 DR InterPro: IPR007087; Znf\_C2H2.  
 DR Pfam: PF00096; Zf-C2H2; 3.  
 DR SMART: SM00355; Znf\_C2H2; 3.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 3.  
 FT NON\_TER 1 1  
 FT NON\_TER 227 227  
 SQ SEQUENCE 227 AA; 25069 MW; C14D4245AF45563C CRC64;  
 Query Match 100.0%; Score 44; DB 2; Length 227;  
 Best Local Similarity 84.6%; Pred. No. 1.3e+02;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAXAAXAAXAAXAA 13  
 DB 58 AAAAAAAAAAAAAA 70  
 RESULT 31  
 Q5K4L2\_9MYRI PRELIMINARY; PRT; 228 AA.  
 AC Q5K4L2;  
 DT 15-FEB-2005, integrated into UniProtKB/TrEMBL.  
 DT 15-FEB-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 8.  
 DE Pax6.2 protein (Fragment).  
 GN Name=Pax6.2;  
 OS Glomeris marginata.  
 OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Pentazonia;  
 CC Glomerida; Glomeridae; Glomeris.  
 OX NCBI\_TaxID=62006;  
 RN [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP PubMed=16351954; DOI=10.1016/j.j.zool.2004.11.003;  
 RT Pp1c N.M.;  
 RT "Duplicated Pax6 genes in Glomeris marginata (Myriapoda: Diplopoda),  
 an arthropod with simple lateral eyes."  
 RL Zoology 108:47-53(2005).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License

CC EMBL: AJ579865; CAE30302.1; -; mRNA.  
 DR SMR; Q5K4L2; 1-103.  
 DR GO: 0005634; C:nucleus; IEA.  
 DR GO: 0003677; F:DNA binding; IEA.  
 DR GO: 0007275; P:development; IEA.  
 DR GO: 0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO: 0006350; P:transcription; IEA.  
 DR InterPro: IPR001523; Paired box N.  
 DR InterPro: IPR011991; Wing\_hlx\_DNA\_bd.  
 DR Pfam: PF00292; PAX; 1.  
 DR PRINTS: PR00027; PAIREDBOX.  
 DR SMART: SM00351; PAX; 1.  
 DR PROSITE: PS51057; PAIRED\_2; 1.  
 KW DNA-binding; Developmental protein; Nuclear protein; Paired box;  
 KW Transcription; Transcription regulation.  
 FT NON\_TER 1 1  
 FT NON\_TER 228 228  
 SQ SEQUENCE 228 AA; 24289 MW; 7CCA50757B2ACDE9 CRC64;  
 Query Match 100.0%; Score 44; DB 2; Length 228;  
 Best Local Similarity 84.6%; Pred. No. 1.3e+02;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAXAAXAAXAAXAA 13  
 DB 129 AAAAAAAAAAAAAA 141  
 RESULT 32  
 Q5BQD5\_9VIRU PRELIMINARY; PRT; 228 AA.  
 AC Q5BQD5;  
 DT 12-APR-2005, integrated into UniProtKB/TrEMBL.  
 DT 12-APR-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 7.  
 DE Coat protein.  
 OS Mint virus X.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flexiviridae;  
 OC Potexvirus; unclassified Potexvirus.  
 OX NCBI\_TaxID=301865;  
 RN [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Ohio;  
 RX PubMed=16096710; DOI=10.1007/s00705-005-0586-x;  
 RA Tzanetakis I.E., Postman J.D., Martin R.R.;  
 RT "Mint virus X: a novel potexvirus associated with symptoms in  
 RT 'Variegata' mint."  
 RL Arch. Virol. 151:143-153(2006).  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 CC EMBL: AY921610; AAX19351.1; -; Genomic RNA.  
 DR GO: 00019028; F:viral capsid; IEA.  
 DR GO: 0005198; F:structural molecule activity; IEA.  
 DR InterPro: IPR000052; Plevir\_coat.  
 DR Pfam: PF00286; Virus\_P-coat; 1.  
 DR PRINTS: PR00232; POTXCARLCOAT.  
 DR PROSITE: PS00418; POTEX\_CARLAVIRUS\_COAT; 1.  
 KW Capsid protein.  
 SQ SEQUENCE 228 AA; 23774 MW; B1D3877B7B11313B CRC64;  
 Query Match 100.0%; Score 44; DB 2; Length 228;  
 Best Local Similarity 84.6%; Pred. No. 1.3e+02;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAXAAXAAXAAXAA 13  
 DB 8 AAAAAAAAAAAAAA 20  
 RESULT 33



DR InterPro: IPR011598; HLH\_DNA\_bd.  
DR Pfam: PF00010; HLH; 1.  
DR SMART: SM00353; HLH; 1.  
DR PROSITE: PS00888; HLH; 1.  
KW Developmental protein; Differentiation; DNA-binding; Neurogenesis;  
KW Nuclear protein.  
FT CHAIN 1 233 Achaete-scute homolog 1.  
FT DOMAIN 129 168 /FTID=PRO\_0000127128.  
FT DNA BIND 118 128 Helix-loop-helix motif.  
FT COMPBIAS 30 44 Basic motif.  
FT COMPBIAS 48 54 Poly-Ala.  
SQ SEQUENCE 233 AA; 24972 MW; 036BDAC8E2D23274 CRC64;  
  
Query Match 100.0%; Score 44; DB 1; Length 233;  
Best Local Similarity 84.6%; Pred. No. 1.3e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAXAAAAAAAXAA 13  
Db 32 AAAAAAAAXAA 44  
  
RESULT 36  
Q7XTV6\_ORYSA PRELIMINARY; PRT; 233 AA.  
AC Q7XTV6;  
DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2004, sequence version 2.  
DT 07-FEB-2006, entry version 14.  
DE OSUNBa0010D21.10 protein.  
GN Name=OSUNBa0010D21.10;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP Clade;  
OC Elnatoidae; Oryzae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22337377; PubMed=12447439; DOI=10.1038/nature01183;  
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,  
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,  
RA Wang Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,  
RA Liu X., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,  
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,  
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,  
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,  
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,  
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,  
RA Han B.;  
RT "Sequence and analysis of rice chromosome 4.";  
RL Nature 420:316-320(2002).  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AL606635; CAD41708.2; -; Genomic DNA.  
DR HSSP: O80337; IGCC.  
DR Gramene; Q7XTV6; -.  
DR GO: GO:0005634; C:nucleus; IEA.  
DR GO: GO:0003700; F:transcription factor activity; IEA.  
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro: IPR001471; TF\_ERF.  
DR Pfam: PF00847; AP2; 1.  
DR ProDom: PD001423; TF\_ERF; 1.  
DR SMART: SM00380; AP2; 1.  
DR PROSITE: PS51032; AP2\_ERF; 1.  
SQ SEQUENCE 233 AA; 25434 MW; 90B3BF38A3A4A419 CRC64;  
  
Query Match 100.0%; Score 44; DB 2; Length 233;  
Best Local Similarity 84.6%; Pred. No. 1.3e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAAXAA 13  
Db 14 AAAAAAAAXAA 26  
  
RESULT 37  
ASCL1\_HUMAN STANDARD; PRT; 236 AA.  
ID ASCL1\_HUMAN STANDARD; PRT; 236 AA.  
AC P50553; Q9BQ30;  
DT 01-OCT-1996, integrated into UniProtKB/Swiss-Prot.  
DT 02-MAY-2002, sequence version 2.  
DT 07-MAR-2006, entry version 46.  
DE Achaete-scute homolog 1 (HASH1).  
GN Name=ASCL1; Synonyms=ASH1;  
OS Homo sapiens (Human).  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA].  
RX MEDLINE=93296195; PubMed=8390674;  
RA Ball D.W., Azzoli C.G., Baylin S.B., Chi D., Dou S., Donis-Keller H.,  
RA Cumarswamy A., Borges M., Nelkin B.D.;  
RT "Identification of a human achaete-scute homolog highly expressed in  
RT neuroendocrine tumors.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:5648-5652(1993).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RX TISSUE=Lung;  
RC MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosk S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: May play a role at early stages of development of  
CC specific neural lineages in most regions of the CNS, and of  
CC several lineages in the PNS. Essential for the generation of  
CC olfactory and autonomic neurons. Activates transcription by  
CC binding to the E box (5'-CANNTG-3').  
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another  
CC bHLH protein. Forms a heterodimer with E12/E47.  
CC -!- SUBCELLULAR LOCATION: Nucleus (Probable).  
CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: L08424; AAS58376.1; -; mRNA.  
DR EMBL: BC001638; -; NOT ANNOTATED CDS; mRNA.  
DR EMBL: BC002341; AAH02341.1; -; mRNA.  
DR EMBL: BC003134; AAH03134.1; -; mRNA.  
DR EMBL: BC004425; AAH04425.1; -; mRNA.  
DR EMBL: BC031299; AAH31299.1; -; mRNA.  
DR PIR: A48279; A48279.  
DR Ensembl: ENSG00000139352; Homo sapiens.



```
DR H-INVD; HIX0010931; -.
DR HGNC; HGNC:738; ASCL1.
DR MIM; 100790; gene.
DR GO; GO:0003700; F:transcription factor activity; NAS.
DR InterPro; IPR001092; HLH basic.
DR InterPro; IPR011598; HLH_DNA_bd.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH; 1.
KW Developmental protein; Differentiation; DNA-binding; Neurogenesis;
KW Nuclear protein; Polymorphism.
FT CHAIN 1 236
FT DOMAIN 132 171
FT DNA_BIND 121 131
FT COMPBIAS 33 47
FT COMPBIAS 51 62
FT VARIANT 158 158
FT CONFLICT 62 62
SQ SEQUENCE 236 AA; 25454 MW; A7D784329305B49A CRC64;

Query Match 100.0%; Score 44; DB 1; Length 236;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
DB 35 AAAAAAAAXAA 47

RESULT 38
Q6QDA4_HUMAN
ID Q6QDA4_HUMAN PRELIMINARY; PRT; 236 AA.
AC Q6QDA4
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE Hypothetical protein.
GN Name=RBM24;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Li H., Zhong G., Zhou G., Wang C., Shen C., Ke R., Li M., Xiao W.,
RA Lin L., Yang S.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AY547318; AAS55633.1; -; mRNA.
DR ENSEMBL; ENSG00000112183; Homo sapiens.
DR HGNC; HGNC:21539; RBM24.
DR GO; GO:0003766; F:nucleic acid binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR012677; a_b_plait_nuc_bd.
DR InterPro; IPR001778; POA_allergen_C.
DR InterPro; IPR000504; RNPI_RNA_bd.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00076; RRM_1; 1.
DR PRINTS; PR00833; POAALLERGEN.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR PROSITE; PS00102; RRM; 1.
SQ SEQUENCE 236 AA; 24740 MW; 9B550B5232E3AA3D CRC64;

Query Match 100.0%; Score 44; DB 2; Length 236;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
DB 35 AAAAAAAAXAA 47

RESULT 39
Q9BX46_HUMAN
ID Q9BX46_HUMAN PRELIMINARY; PRT; 236 AA.
AC Q9BX46;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 28.
DE OTTHUMP00000016066.
GN ORFNames=RP1-259A10.1-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Babbage A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AL136305; CAC36889.1; -; Genomic DNA.
DR HSSP; P09651; IL3K.
DR ENSEMBL; ENSG00000112183; Homo sapiens.
DR GO; GO:0003766; F:nucleic acid binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR012677; a_b_plait_nuc_bd.
DR InterPro; IPR001778; POA_allergen_C.
DR InterPro; IPR000504; RNPI_RNA_bd.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00076; RRM_1; 1.
DR PRINTS; PR00833; POAALLERGEN.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR PROSITE; PS00102; RRM; 1.
SQ SEQUENCE 236 AA; 24776 MW; 1CFB5AEBD4E3AA24 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 236;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
DB 209 AAAAAAAAXAA 221

RESULT 40
Q6CNB7_KLUJLA
ID Q6CNB7_KLUJLA PRELIMINARY; PRT; 244 AA.
AC Q6CNB7;
DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2004, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Similarities with sp|Q9VAC5 Aeropyrum pernix Hypothetical protein
DE APE2014.
GN OrderedLocusNames=KLJA0E13816g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
```

```
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
DB 209 AAAAAAAAXAA 221

RESULT 39
Q9BX46_HUMAN
ID Q9BX46_HUMAN PRELIMINARY; PRT; 236 AA.
AC Q9BX46;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 28.
DE OTTHUMP00000016066.
GN ORFNames=RP1-259A10.1-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Babbage A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AL136305; CAC36889.1; -; Genomic DNA.
DR HSSP; P09651; IL3K.
DR ENSEMBL; ENSG00000112183; Homo sapiens.
DR GO; GO:0003766; F:nucleic acid binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR InterPro; IPR012677; a_b_plait_nuc_bd.
DR InterPro; IPR001778; POA_allergen_C.
DR InterPro; IPR000504; RNPI_RNA_bd.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00076; RRM_1; 1.
DR PRINTS; PR00833; POAALLERGEN.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN_1.
DR PROSITE; PS00102; RRM; 1.
SQ SEQUENCE 236 AA; 24776 MW; 1CFB5AEBD4E3AA24 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 236;
Best Local Similarity 84.6%; Pred. No. 1.3e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13
DB 209 AAAAAAAAXAA 221

RESULT 40
Q6CNB7_KLUJLA
ID Q6CNB7_KLUJLA PRELIMINARY; PRT; 244 AA.
AC Q6CNB7;
DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2004, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Similarities with sp|Q9VAC5 Aeropyrum pernix Hypothetical protein
DE APE2014.
GN OrderedLocusNames=KLJA0E13816g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
```

RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., V.,  
RA Goffard N., Franchin L., Aigle M., Anthouard V., Babour A., Barbe V.,  
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
RA Boisrame A., Boyer J., Cattolico L., Confanioulet F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
RA Swennen D., Tekaita F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
RT Wincker P., Souciet J.-L.;  
RL "Genome evolution in yeasts.";  
RL Nature 430:35-44 (2004).

CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CR382125; CAG99659.1; -; Genomic DNA.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 244 AA; 24930 MW; CC83CACE35423EE2 CRC64;

Query Match 100.0%; Score 44; DB 2; Length 244;  
Best Local Similarity 84.6%; Pred. No. 1.4e+02;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AAXAAAAAXAA 13  
| | | | | | | | | |  
Db 144 AAAAAAAAAAA 156

Search completed: September 9, 2006, 23:00:22  
Job time : 120.139 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:48:41 ; Search time 14.8101 Seconds  
(without alignments)  
84.457 Million cell updates/sec

Title: US-10-617-568-5  
Perfect score: 54  
Sequence: 1 ASMSAASAAASMAA 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database : PIR\_80:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	77.8	328	2 AB1781	hypothetical prote
2	40	74.1	510	2 A32380	nuclear protein fk
3	39	72.2	470	2 T22785	hypothetical prote
4	39	72.2	799	2 JH0797	castor protein - f
5	38	70.4	328	2 AC1405	hypothetical prote
6	38	70.4	508	2 S59870	fork head domain p
7	37	68.5	375	2 AC0484	probable regulator
8	36	66.7	258	2 D97395	hypothetical 31.3K
9	36	66.7	268	2 B61615	fibron heavy chai
10	36	66.7	313	2 AF2613	tetrapyrrole methy
11	36	66.7	329	2 AC1718	ferrichrome ABC tr
12	36	66.7	329	2 AG1347	ferrichrome ABC tr
13	36	66.7	407	2 T19155	hypothetical prote
14	36	66.7	420	2 T36193	probable salicylat
15	36	66.7	474	2 I38240	transcription fact
16	36	66.7	509	2 T02940	O-succinylhomoseri
17	36	66.7	509	2 T02942	O-succinylhomoseri
18	36	66.7	537	2 A55929	zinc finger protei
19	36	66.7	686	1 VGBENA	glycoprotein H pre
20	36	66.7	686	1 VGBEPK	glycoprotein H pre
21	36	66.7	686	2 S15478	glycoprotein H - s
22	36	66.7	2023	2 T13154	polycarb protein e
23	36	66.7	5069	2 T17464	rifamycin polyketi
24	35	64.8	308	2 AB3914	ABC transporter (p
25	35	64.8	314	1 JC4210	3-hydroxyacyl-CoA
26	35	64.8	339	2 T26328	hypothetical prote
27	35	64.8	436	2 G81655	conserved hypothet
28	35	64.8	439	2 E71497	hypothetical prote
29	35	64.8	535	2 H83324	probable chemotaxi

30	35	64.8	644	2 S39356	transcription fact
31	35	64.8	646	2 T42296	hypothetical prote
32	35	64.8	753	2 A27041	tyrosine kinase-re
33	35	64.8	771	2 S35681	ESG protein - mous
34	35	64.8	1122	2 G64887	probable tail fibe
35	35	64.8	1520	2 AF3008	polyketide synthas
36	35	64.8	1520	2 G98275	hypothetical prote
37	35	64.8	2957	2 T33152	hypothetical prote
38	34	63.0	144	1 SXAD12	hexon-associated p
39	34	63.0	169	2 T03033	probable cytochrom
40	34	63.0	211	2 T46497	hypothetical prote
41	34	63.0	250	2 H83213	hypothetical prote
42	34	63.0	257	2 B69213	protein-export mem
43	34	63.0	308	2 A44496	transcription repr
44	34	63.0	333	2 A39065	homeotic protein E
45	34	63.0	475	2 A43915	homeotic protein e
46	34	63.0	485	2 S36184	translation elonga
47	34	63.0	524	2 A86356	hypothetical prote
48	34	63.0	557	2 S12359	nicotinic acetylch
49	34	63.0	667	2 S63587	gene pac protein
50	34	63.0	712	2 H87509	conserved hypothet
51	34	63.0	722	2 S57246	ventral nervous sy
52	34	63.0	724	2 T47149	hypothetical prote
53	34	63.0	741	2 B49555	enhancer of split
54	34	63.0	772	2 D56695	transducin-like en
55	34	63.0	956	1 QQBEK2	ULI05 protein - hu
56	34	63.0	2946	2 T00867	hypothetical prote
57	33	61.1	54	2 T29665	hypothetical prote
58	33	61.1	85	1 FDFL4W	antifreeze protein
59	33	61.1	91	2 A22592	antifreeze protein
60	33	61.1	98	2 C70638	hypothetical prote
61	33	61.1	106	2 JN0428	ribonuclease Tn1 (
62	33	61.1	126	2 E90726	probable homeobox
63	33	61.1	126	2 F85577	probable homeobox
64	33	61.1	126	2 A64811	YbGS protein precu
65	33	61.1	128	2 T30714	hypothetical prote
66	33	61.1	234	2 A11852	hypothetical prote
67	33	61.1	364	2 I48188	gene NKx6.1 protei
68	33	61.1	391	2 D70922	probable PPE prote
69	33	61.1	400	2 T45319	hypothetical prote
70	33	61.1	401	2 E70646	probable fadE23 pr
71	33	61.1	407	2 S22586	homeotic protein E
72	33	61.1	414	2 D96524	protein Tln15.3 (i
73	33	61.1	416	2 S12541	evx1 protein - mur
74	33	61.1	418	2 E47056	nickel-cobalt resi
75	33	61.1	443	2 T48708	hypothetical prote
76	33	61.1	448	2 S17370	DNA-binding protei
77	33	61.1	535	2 T04402	probable glutamyl-
78	33	61.1	577	2 AD1440	hypothetical prote
79	33	61.1	600	2 S07638	spore coat protein
80	33	61.1	601	2 S33377	P63 protein - huma
81	33	61.1	604	2 A33369	homeotic protein B
82	33	61.1	606	2 S13367	Om(1D) protein - f
83	33	61.1	627	2 T02610	probable YME1 ATP-
84	33	61.1	630	2 AE3463	hypothetical membr
85	33	61.1	638	2 D85435	BEI1-like homeobox
86	33	61.1	640	2 A41726	homeotic protein B
87	33	61.1	642	2 S27806	homeotic protein B
88	33	61.1	660	2 C86553	heat shock protein
89	33	61.1	660	2 S16159	dnak-type molecula
90	33	61.1	664	2 S53037	PLB1 protein - yea
91	33	61.1	723	2 T14605	probable cell divi
92	33	61.1	807	2 A53225	ecdysone-induced p
93	33	61.1	827	2 S50714	DNA-binding protei
94	33	61.1	829	2 A34692	ecdysone-induced p
95	33	61.1	832	2 A31246	neurogenic protein
96	33	61.1	833	2 S19087	gene Delta protein
97	33	61.1	836	2 S49940	cell division cont
98	33	61.1	880	2 S00670	neurogenic repetit
99	33	61.1	883	2 S04722	puff 74E protein -
100	33	61.1	971	2 B90835	probable tail fibe

## ALIGNMENTS

```
RESULT 1
AB1781
hypothetical protein lin2792 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AB1781
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1781
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <GLA>
A:Cross-references: UNIPROT:Q927J6; UNIPARC:UPI00000CC998; GB:AL592022; PIDN:CAC98018.1;
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin2792

Query Match 77.8%; Score 42; DB 2; Length 328;
Best Local Similarity 76.9%; Pred. No. 10;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
Db 313 AASMAASAAASMAA 325
|:|||||:|

RESULT 2
A32380
nuclear protein fkh - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 08-Sep-1989 #sequence_revision 08-Sep-1989 #text_change 31-Dec-2004
C:Accession: A32380
R:Weigel, D.; Juergens, G.; Kuettnr, F.; Seifert, E.; Jaecle, H.
Cell 57, 645-658, 1989
A:Title: The homeotic gene fork head encodes a nuclear protein and is expressed in the b
A:Reference number: A32380; MUID:89249328; PMID:2566386
A:Accession: A32380
A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-510 <WEI>
A:Cross-references: UNIPROT:P14734; UNIPARC:UPI000012457F; GB:J03177; NID:g157433; PIDN:
C:Genetics:
A:Gene: FlyBase:fkh
A:Cross-references: FlyBase:FBgn0000659
C:Keywords: DNA binding; transcription regulation
F:210-301/Domain: fork head DNA-binding domain homology <FHD>

Query Match 74.1%; Score 40; DB 2; Length 510;
Best Local Similarity 69.2%; Pred. No. 32;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
Db 101 ASMSASMSASMAA 113
|:|||||:|

RESULT 3
T22785
hypothetical protein F56G4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22785
R:Lennard, N.
A: submitted to the EMBL Data Library, November 1996
```

```
A:Reference number: Z19615
A:Accession: T22785
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-470 <WIL>
A:Cross-references: UNIPROT:Q9XUY1; UNIPARC:UPI000007519F; EMBL:Z81552; PIDN:CAB04486.1;
A:Experimental source: clone F56G4
C:Genetics:
A:Gene: CESP:F56G4.4
A:Map position: 1
A:Introns: 17/3; 152/1; 241/3; 339/1; 449/3
C:Superfamily: hypothetical protein F56G4.4; WW repeat homology
F:177-216/Domain: WW repeat homology <WWR>

Query Match 72.2%; Score 39; DB 2; Length 470;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMA 12
Db 66 AKMEASAAASMA 77
|:|||||:|

RESULT 4
JH0797
castor protein - fruit fly (Drosophila sp.)
C:Species: Drosophila sp.
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C:Accession: JH0797
R:Mellerick, D.M.; Kassis, J.A.; Zhang, S.D.; Odenwald, W.F.
Neuron 9, 789-803, 1992
A:Title: Castor encodes a novel zinc finger protein required for the development of a su
A:Reference number: JH0797; MUID:93040222; PMID:1418995
A:Accession: JH0797
A:Molecule type: DNA
A:Residues: 1-799 <MEL>
A:Cross-references: UNIPROT:Q7M3M8; UNIPARC:UPI000017BEED; GB:L04487
C:Comment: This protein is required for the development of embryonic CNS neurons.
C:Genetics:
A:Gene: FlyBase:cas
A:Cross-references: FlyBase:FBgn0004878
A:Introns: 284/3; 515/3; 660/3
C:Keywords: metal binding; zinc finger
F:236-242/Region: acidic
F:363-421/Region: zinc finger
F:422-482/Region: zinc finger
F:483-541/Region: zinc finger
F:542-600/Region: zinc finger

Query Match 72.2%; Score 39; DB 2; Length 799;
Best Local Similarity 69.2%; Pred. No. 69;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
Db 758 AALSAAASAAAAA 770
|:|||||:|

RESULT 5
AC1405
hypothetical protein lmo2643 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1405
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1405
```

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-328 <GLA>
A:Cross-references: UNIPROT:Q8Y431; UNIPARC:UPI00000556C4; GB:NC_003210; PIDN:CAD00721.1
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2643

Query Match      70.4%; Score 38; DB 2; Length 328;
Best Local Similarity 69.2%; Pred. No. 44;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
   ||:|||||:|
Db 313 AAVSAAASAAASAA 325

RESULT 6
S59870
fork head domain protein crocodile - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 05-Oct-2004
C:Accession: S59870; A46178
R:Haacker, U.; Kaufmann, E.; Hartmann, C.; Juergens, G.; Knoechel, W.; Jaeckle, H.
EMBO J. 14, 5306-5317, 1995
A:Title: The Drosophila fork head domain protein crocodile is required for the establish
A:Reference number: S59870; MUID:96080166; PMID:7489720
A:Accession: S59870
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-508 <HAB>
A:Cross-references: UNIPROT:P32027; UNIPARC:UPI0000001B0F
R:Haacker, U.; Grossniklaus, U.; Gehring, W.J.; Jackle, H.
Proc. Natl. Acad. Sci. U.S.A. 89, 8754-8758, 1992
A:Title: Developmentally regulated Drosophila gene family encoding the fork head domain.
A:Reference number: A46178; MUID:92409595; PMID:1356269
A:Accession: A46178
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 55-182 <HAC>
A:Cross-references: UNIPARC:UPI000016BBEC; GB:M96440; NID:gl57425; PIDN:AAF02177.1; PID:
A>Note: sequence extracted from NCBI backbone (NCBIP:114222)
C:Genetics:
A:Gene: croc
A:Cross-references: FlyBase:FBgn0014143
P:70-161/Domain: fork head DNA-binding domain homology <FHD>

Query Match      70.4%; Score 38; DB 2; Length 508;
Best Local Similarity 69.2%; Pred. No. 66;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
   ||:|||||:|
Db 449 ASVAASAAASAAAA 461

RESULT 7
AC0484
probable regulatory protein YPO3978 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC0484
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0484
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <KUR>
A:Cross-references: UNIPROT:Q8ZA42; UNIPARC:UPI000000DC955; GB:AL590842; PIDN:CAC93439.1;

```

```

C:Genetics:
A:Gene: YPO3978
C:Superfamily: hypothetical protein HI0093

Query Match      68.5%; Score 37; DB 2; Length 375;
Best Local Similarity 90.0%; Pred. No. 72;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 SAASAAASMAA 13
   |||||
Db 145 SAASVASMAA 154

RESULT 8
D97395
hypothetical 31.3K protein in agaI-mtr intergenic region (f286) [imported] - Agrobacter
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: D97395
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: D97395
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <KUR>
A:Cross-references: UNIPROT:Q8ULJ3; UNIPARC:UPI0000164347; GB:AE007869; PIDN:AAK86117.1
C:Genetics:
A:Gene: AGR_C523
A:Map position: circular chromosome
A:Superfamily: methyltransferase, YraL type

Query Match      66.7%; Score 36; DB 2; Length 258;
Best Local Similarity 69.2%; Pred. No. 74;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
   ||:|||||:|
Db 223 ATMSAAKAAAGEAA 235

RESULT 9
B61615
fibroin heavy chain PG-2' - greater wax moth (fragments)
C:Species: Galleria mellonella (greater wax moth)
C>Date: 18-Jul-1995 #sequence_revision 28-Jul-1995 #text_change 09-Jul-2004
C:Accession: B61615
R:Zurovec, M.; Sehna, F.; Scheller, K.; Kumatan, A.K.
Insect Biochem. Mol. Biol. 22, 55-67, 1992
A:Title: Silk gland specific cDNAs from Galleria mellonella L.
A:Reference number: A61615
A:Accession: B61615
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-268 <ZUR>
A:Cross-references: UNIPROT:Q7M468; UNIPARC:UPI000017CAA4

Query Match      66.7%; Score 36; DB 2; Length 268;
Best Local Similarity 76.9%; Pred. No. 76;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
   |||||
Db 98 AGSSAASAAASGAA 110

RESULT 10
AF2613
tetrapyrrole methylase family protein [imported] - Agrobacterium tumefaciens (strain C5)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004

```

C;Accession: AF2613  
R;Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; UID:21608550; PMID:11743193  
A;Accession: AF2613  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-313 <GUA>  
A;Cross-references: UNIPROT:Q8UIJ3; UNIPARC:UPI000000DI77D; GB:AE008688; PIDN:AAL41324.1  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: Atu0302  
A;Map position: circular chromosome  
C;Superfamily: methyltransferase, YraI type

Query Match 66.7%; Score 36; DB 2; Length 313;  
Best Local Similarity 69.2%; Pred. No. 88;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
Db 278 ATWSAAKAGEAA 290  
:|||||:|

RESULT 11  
AC1718  
ferrichrome ABC transporter (permease) homolog lin2287 [imported] - Listeria innocua (st  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
A;Accession: AC1718  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; UID:21537279; PMID:11679669  
A;Accession: AC1718  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-329 <GUA>  
A;Cross-references: UNIPROT:Q929I8; UNIPARC:UPI00000CC7FB; GB:AL592022; PIDN:CAC97515.1  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin2287

Query Match 66.7%; Score 36; DB 2; Length 329;  
Best Local Similarity 66.7%; Pred. No. 92;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMA 12  
Db 169 SSQSASSAASLA 180  
:|||||:|

RESULT 12  
AG1347  
ferrichrome ABC transporter (permease) homolog lmc2183 [imported] - Listeria monocytogen  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
A;Accession: AG1347  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; UID:21537279; PMID:11679669  
A;Accession: AG1347  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-329 <GUA>  
A;Cross-references: UNIPROT:Q8Y586; UNIPARC:UPI000005571F; GB:NC\_003210; PIDN:CAD00261.1  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: lmo2183

Query Match 66.7%; Score 36; DB 2; Length 329;  
Best Local Similarity 66.7%; Pred. No. 92;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMA 12  
Db 169 SSQSASSAASLA 180  
:|||||:|

RESULT 13  
TI9155  
hypothetical protein C09G9.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
A;Accession: TI9155; TI9510  
R;Berks, M.  
submitted to the EMBL Data Library, September 1995  
A;Reference number: Z19081  
A;Accession: TI9155  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-407 <WIL>  
A;Cross-references: UNIPROT:Q17877; UNIPARC:UPI00000754B2; EMBL:Z54235; PIDN:CAA90977.1  
A;Experimental source: clone C09G9  
R;Matthews, P.  
submitted to the EMBL Data Library, September 1995  
A;Reference number: Z19133  
A;Accession: TI9510  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-407 <W12>  
A;Cross-references: UNIPARC:UPI00000754B2; EMBL:Z54236; PIDN:CAA90986.1; GSPDB:GN00022;  
A;Experimental source: clone C27B7  
C;Genetics:  
A;Gene: CESP:C09G9.6  
A;Map position: 4  
A;Introns: 13/2; 69/2; 146/2; 195/3; 273/1  
C;Superfamily: Caenorhabditis elegans hypothetical protein ZC513.6

Query Match 66.7%; Score 36; DB 2; Length 407;  
Best Local Similarity 61.5%; Pred. No. 11e+02;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
Db 347 SSLSAASAAAAA 359  
:|||||:|

RESULT 14  
T36193  
probable salicylate hydroxylase - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
A;Accession: T36193  
R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrall, B.G.; Rajandream, M  
submitted to the EMBL Data Library, March 1999  
A;Reference number: Z21600  
A;Accession: T36193  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-420 <SAU>  
A;Cross-references: UNIPROT:Q9Z4Y6; UNIPARC:UPI000000DAF36; EMBL:AL035707; PIDN:CAB38889

A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCODB:SCB29.14c  
C:Superfamily: tetracycline 6-hydroxylase

Query Match 66.7%; Score 36; DB 2; Length 420;  
Best Local Similarity 69.2%; Pred. No. 1.1e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
|:|:|:|:|:|:|  
DB 327 ASVSASASASAPA 339

RESULT 15  
T02940  
O-succinylhomoserine (thiol)-lyase (EC 4.2.99.9) 1 - maize  
N:Alternate names: sex-determining region Y box 4 protein  
C:Species: Homo sapiens (man)  
C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 05-Oct-2004  
A:Accession: I38240; S22938; S21479; S31724  
R:Farr, C.J.; Easty, D.J.; Ragoussis, J.; Collignon, J.; Lovell-Badge, R.; Goodfellow, H.  
Mamm. Genome 4, 577-584, 1993  
A:Title: Characterization and mapping of the human SOX4 gene.  
A:Reference number: I38240; MUID:94093204; PMID:8268656  
A:Accession: I38240  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-474 <RES>  
A:Cross-references: UNIPROT:Q06945; UNIPARC:UPI0000047FA0; EMBL:X70683; NID:g36552; PIDN:Denny, P.; Swift, S.; Brand, N.; Dabhadre, N.; Barton, P.; Ashworth, A.  
Nucleic Acids Res. 20, 2887, 1992  
A:Title: A conserved family of genes related to the testis determining gene, SRY.  
A:Reference number: S22935; MUID:92310593; PMID:1614875  
A:Accession: S22938  
A:Molecule type: mRNA  
A:Residues: 70,'P',72-123 <DEN>  
A:Cross-references: UNIPARC:UPI000016B04E; EMBL:X65661; NID:g36551; PIDN:CAA46612.1; PID  
C:Genetics:  
A:Gene: GDB:SOX4  
A:Cross-references: GDB:250365; OMIM:184430  
A:Map position: 6p22.2-6p22.16p23-6p23  
C:Keywords: DNA binding; transcription regulation  
F:56-131/Domain: HMG box homology <HMG1>

Query Match 66.7%; Score 36; DB 2; Length 474;  
Best Local Similarity 69.2%; Pred. No. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
|:|:|:|:|:|:|  
DB 267 ASASASASASASA 279

RESULT 16  
T02940  
O-succinylhomoserine (thiol)-lyase (EC 4.2.99.9) 1 - maize  
N:Alternate names: cystathionine gamma-synthase  
C:Species: Zea mays (maize)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 05-Oct-2004  
A:Accession: T02940  
R:Locke, M.E.H.; Guida, A.D.; Falco, S.C.  
submitted to the EMBL Data Library, June 1997  
A:Description: Deregulation of the methionine biosynthetic pathway in corn seeds.  
A:Reference number: Z14785  
A:Accession: T02940  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-509 <LOC>  
A:Cross-references: UNIPROT:Q04981; UNIPARC:UPI00000A35C3; EMBL:AF007785; NID:g2198850;  
A:Experimental source: strain H99  
C:Genetics:  
A:Gene: CGSI

C:Superfamily: cystathionine gamma-synthase  
C:Keywords: carbon-oxygen lyase

Query Match 66.7%; Score 36; DB 2; Length 509;  
Best Local Similarity 69.2%; Pred. No. 1.4e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
|:|:|:|:|:|:|  
DB 83 AAASASASASAAA 95

RESULT 17  
T02942  
O-succinylhomoserine (thiol)-lyase (EC 4.2.99.9) 1 - maize  
N:Alternate names: cystathionine gamma-synthase  
C:Species: Zea mays (maize)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 05-Oct-2004  
A:Accession: T02942  
R:Locke, M.E.H.; Guida, A.D.; Falco, S.C.  
submitted to the EMBL Data Library, June 1997  
A:Description: Deregulation of the methionine biosynthetic pathway in corn seeds.  
A:Reference number: Z14785  
A:Accession: T02942  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-509 <LOC>  
A:Cross-references: UNIPROT:Q04982; UNIPARC:UPI00000A2615; EMBL:AF007786; NID:g2198852;  
A:Experimental source: strain Missouri 17  
C:Genetics:  
A:Gene: CGSI  
A:Introns: 129/1; 163/3; 188/2; 244/3; 264/3; 332/3; 361/3; 397/3; 429/3; 469/2  
C:Superfamily: cystathionine gamma-synthase  
C:Keywords: carbon-oxygen lyase

Query Match 66.7%; Score 36; DB 2; Length 509;  
Best Local Similarity 69.2%; Pred. No. 1.4e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
|:|:|:|:|:|:|  
DB 83 AAASASASASAAA 95

RESULT 18  
AS5929  
zinc finger protein noca - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Jul-2004  
A:Accession: AS5929  
R:Cheah, P.Y.; Meng, Y.B.; Yang, X.; Kimbrell, D.; Ashburner, M.; Chia, W.  
Mol. Cell. Biol. 14, 1487-1499, 1994  
A:Title: The Drosophila 1(2)35Ba/noca gene encodes a putative Zn finger protein involved  
A:Reference number: AS5929; MUID:94119100; PMID:8289824  
A:Accession: AS5929  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-537 <CHE>  
A:Cross-references: UNIPROT:Q24423; UNIPARC:UPI000007B356; GB:L14009; NID:g431289; PID:  
C:Genetics:  
A:Gene: FlyBase:noc  
A:Cross-references: FlyBase:FBgn0005771

Query Match 66.7%; Score 36; DB 2; Length 537;  
Best Local Similarity 61.5%; Pred. No. 1.4e+02;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
|:|:|:|:|:|:|  
DB 232 SSMAASASASAAA 244

RESULT 19

## VGBENA

glycoprotein H precursor - suid herpesvirus 1 (strain NIA-3)  
C:Species: suid herpesvirus 1  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 09-Jul-2004  
C:Accession: A42000; S16737  
R:Peeters, B.; de Wind, N.; Broer, R.; Gielkens, A.; Moormann, R.  
J. Virol. 66, 3888-3892, 1992  
A:Title: Glycoprotein H of pseudorabies virus is essential for entry and cell-to-cell spread  
A:Reference number: A42000; PMID:92260665; PMID:1316488  
A:Accession: A42000  
A:Molecule type: DNA  
A:Residues: 1-686 <PE>  
A:Cross-references: UNIPROT:Q00660; UNIPARC:UPI0000138710; GB:X61696; NID:961352; PIDN:C  
C:Superfamily: herpesvirus glycoprotein H  
C:Keywords: glycoprotein; transmembrane protein  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-686/Product: glycoprotein H #status predicted <GPH>  
F:647-667/Domain: transmembrane #status predicted <TMN>  
F:77,162,542,604,627/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 66.7%; Score 36; DB 1; Length 686;  
Best Local Similarity 61.5%; Pred. No. 1.8e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
Db 245 AQLSAHAHAALAA 257

## RESULT 20

VGBEPK  
glycoprotein H precursor - suid herpesvirus 1 (strain Ka)  
C:Species: suid herpesvirus 1  
C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C:Accession: A39990  
R:Klupp, B.G.; Mettenleiter, T.C.  
Virology 182, 732-741, 1991  
A:Title: Sequence and expression of the glycoprotein gH gene of pseudorabies virus.  
A:Reference number: A39990; MUID:91220723; PMID:1850925  
A:Accession: A39990  
A:Molecule type: DNA  
A:Residues: 1-686 <KLU>  
A:Cross-references: UNIPROT:P27416; UNIPARC:UPI000013870F; GB:M61196; NID:G334058; PIDN:  
C:Superfamily: herpesvirus glycoprotein H  
C:Keywords: glycoprotein; transmembrane protein  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-686/Product: glycoprotein H #status predicted <GPH>  
F:647-667/Domain: transmembrane #status predicted <TMN>  
F:77,162,542,604,627/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 66.7%; Score 36; DB 1; Length 686;  
Best Local Similarity 61.5%; Pred. No. 1.8e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
Db 245 AQLSAHAHAALAA 257

## RESULT 21

S15478  
glycoprotein H - suid herpesvirus 1  
C:Species: suid herpesvirus 1  
C:Date: 09-Jun-1994 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
C:Accession: S15478  
R:Meyer, A.L.; Petrovskis, E.A.; Thomsen, D.R.; Post, L.E.  
submitted to the EMBL Data Library, April 1991  
A:Description: Cloning and sequence of a pseudorabies virus gene homologous to glycoprotein H  
A:Reference number: S15478  
A:Accession: S15478  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-686 <MEY>

A:Cross-references: UNIPROT:P27593; UNIPARC:UPI0000138711; EMBL:X58868; NID:G59965; PIDN:  
A:Note: the source is given as pseudorabies virus  
C:Superfamily: herpesvirus glycoprotein H  
C:Keywords: glycoprotein; transmembrane protein  
Query Match 66.7%; Score 36; DB 2; Length 686;  
Best Local Similarity 61.5%; Pred. No. 1.8e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 ASMSAASAAASMAA 13  
Db 245 AQLSAHAHAALAA 257

## RESULT 22

T13154  
polycomb protein enhancer - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: T13154  
R:Stankunas, K.; Berger, J.; Ruse, C.; Sinclair, D.A.; Randazzo, F.; Brock, H.W.  
Development 125, 4055-4066, 1998  
A:Title: The enhancer of polycomb gene of Drosophila encodes a chromatin protein conserv  
A:Reference number: Z17611; MUID:98407961; PMID:9735366  
A:Accession: T13154  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-2023 <STA>  
A:Cross-references: UNIPROT:O96542; UNIPARC:UPI0000083EF3; EMBL:AF079764; NID:G3757889;  
A:Experimental source: imaginal disc  
C:Genetics:  
A:Gene: E(Pc)  
A:Cross-references: FlyBase:FBgn0000581  
A:Map position: 2

Query Match 66.7%; Score 36; DB 2; Length 2023;  
Best Local Similarity 69.2%; Pred. No. 4.8e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
Db 816 ASTSAHAHAALAA 828

## RESULT 23

T17464  
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei  
C:Species: Amycolatopsis mediterranei  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: T17464  
R:Schupp, T.  
submitted to the EMBL Data Library, December 1997  
A:Reference number: Z18802  
A:Accession: T17464  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-5069 <SCH>  
A:Cross-references: UNIPROT:O52789; UNIPARC:UPI0000055B02; EMBL:AJ223012; NID:e1227119;  
A:Experimental source: strain LBG A3136  
C:Keywords: carrier protein  
F:1631-1702/Domain: acyl carrier protein homology <ACP1>  
F:3238-3309/Domain: acyl carrier protein homology <ACP2>  
F:4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 66.7%; Score 36; DB 2; Length 5069;  
Best Local Similarity 72.7%; Pred. No. 1.1e+03;  
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 3 MSAAASAAASMAA 13  
Db 490 VSNATTAALAA 500



```
RESULT 24
A83914
ABC transporter (permease) BH2113 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C:Accession: A83914
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A>Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83914
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <STO>
A:Cross-references: UNIPROT:Q9K922; UNIPARC:UPI00000D7427; GB:BA000004; NID
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2113
C:Superfamily: maltose transport protein malG

Query Match 64.8%; Score 35; DB 2; Length 308;
Best Local Similarity 69.2%; Pred. No. 1.3e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13
||||| |||||
Db 247 ASMSSGSAASAPA 259

RESULT 25
JC4210
3-hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35), short chain-specific, precursor - mouse
N:Alternate names: RAD-38 protein
C:Species: Mus musculus (house mouse)
C>Date: 04-Oct-1995 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C:Accession: JC4210
R:Nomura, M.; Takiwara, Y.; Shimada, K.
Gene 160, 309-310, 1995
A>Title: Isolation of a cDNA clone encoding mouse 3-hydroxyacyl CoA dehydrogenase.
A:Reference number: JC4210; MUID:95369712; PMID:7642117
A:Accession: JC4210
A:Molecule type: mRNA
A:Residues: 1-314 <NOM>
A:Cross-references: UNIPROT:Q61425; UNIPARC:UPI0000171F45; DDBJ:D29639
A:Experimental source: embryonal carcinoma F9 cells
C:Comment: This enzyme plays a role in the mitochondrial beta-oxidation of short chain f
1-CoA's with concomitant reduction of NAD to NADH and exerts its highest activity toward
C:Superfamily: 3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA dehydrogenase homology
C:Keywords: fatty acid beta-oxidation; homodimer; mitochondrion; NAD; oxidoreductase
F:1-12/Domain: transit peptide (mitochondrion) #status predicted <TP>
F:13-314/Product: 3-hydroxyacyl-CoA dehydrogenase, short chain-specific #status predicted
F:27-314/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>
F:29-57/Region: beta-alpha-beta NAD nucleotide-binding fold

Query Match 64.8%; Score 35; DB 1; Length 314;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SMSAASASMAA 13
||||| |||||
Db 11 SMSSSSASAAA 22

RESULT 26
T26328
hypothetical protein W10D5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26328
R:Kershaw, J.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z20200
A:Accession: T26328
```

```
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-339 <WIL>
A:Cross-references: UNIPARC:UPI000017BC34; EMBL:Z79758; NID:e1062260; PIDN:CAB02131.1;
A:Experimental source: clone W10D5
C:Genetics:
A:Gene: CESP:W10D5.1
A:Map position: 1
A:Introns: 18/3; 56/2; 195/2; 269/1

Query Match 64.8%; Score 35; DB 2; Length 339;
Best Local Similarity 61.5%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13
||| :||| :|||
Db 131 ASASASAAVA 143

RESULT 27
G81655
conserved hypothetical protein TC0868 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: G81655
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: G81655
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-436 <TET>
A:Cross-references: UNIPROT:Q9PJG1; UNIPARC:UPI0000057ABA; GB:AE002353; GB:AE002160; NI
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0868
C:Superfamily: conserved hypothetical protein TC0868

Query Match 64.8%; Score 35; DB 2; Length 436;
Best Local Similarity 69.2%; Pred. No. 1.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13
|: ||||| ||
Db 268 AASGAASASAA 280

RESULT 28
E71497
hypothetical protein CT579 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C:Accession: E71497
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell
Science 282, 754-759, 1998
A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tra
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: E71497
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-439 <ARN>
A:Cross-references: UNIPROT:O84583; UNIPARC:UPI0000139ACF; GB:AE001328; GB:AE001273; NI
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT579
C:Superfamily: conserved hypothetical protein TC0868

Query Match 64.8%; Score 35; DB 2; Length 439;
Best Local Similarity 69.2%; Pred. No. 1.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 ASMSAASAAASMAA 13  
|: |||||  
Db 271 ATSGAASAAASNA 283

## RESULT 29

H83324  
Probable chemotaxis transducer PA2573 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: H83324  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: H83324  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-535 <STO>  
A:Cross-references: UNIPROT:Q910R3; UNIPARC:UPI00000C5747; GB:AE004685; GB:AE004091; NID:10984043  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2573

Query Match 64.8%; Score 35; DB 2; Length 535;  
Best Local Similarity 69.2%; Pred. No. 2.1e+02;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13  
|: |||||  
Db 285 ASQQADSATSMAA 297

## RESULT 30

S39356  
transcription factor btd - fruit fly (Drosophila sp.)  
C:Species: Drosophila sp.  
C:Date: 18-Feb-1994 #sequence\_revision 01-Dec-1995 #text\_change 07-May-1999  
C:Accession: S39356  
R:Wimmer, E.A.; Jaecckle, H.; Pfeifle, C.; Cohen, S.M.  
Nature 366, 690-694, 1993  
A:Title: A Drosophila homologue of human Sp1 is a head-specific segmentation gene.  
A:Reference number: S39356; MUID:94081952; PMID:8259212  
A:Accession: S39356  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <WIM>  
A:Cross-references: UNIPARC:UPI0000124C17; EMBL:229361; NID:g441283; PID:g441284  
C:Genetics:  
A:Gene: FlyBase:btd  
A:Cross-references: FlyBase:FBgn0000233  
A:Introns: 245/2

Query Match 64.8%; Score 35; DB 2; Length 644;  
Best Local Similarity 75.0%; Pred. No. 2.4e+02;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SMSAASAAASMAA 13  
|: |||||  
Db 202 SSSAASAAAAA 213

## RESULT 31

T42296  
hypothetical protein - phage SPPI  
C:Species: phage SPPI  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T42296  
R:Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, T.A.  
Gene 204, 201-212, 1997  
A:Title: The complete nucleotide sequence and functional organization of Bacillus subtilis

A:Reference number: 222137; MUID:98094274; PMID:9434185  
A:Accession: T42296  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-646 <ALO>  
A:Cross-references: UNIPROT:O48456; UNIPARC:UPI000009B4CA; EMBL:X97918; PIDN:CAA66557.1

Query Match 64.8%; Score 35; DB 2; Length 646;  
Best Local Similarity 61.5%; Pred. No. 2.5e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13  
|: |||||  
Db 202 AAQSAANAVKAA 214

## RESULT 32

A27041  
tyrosine kinase-related protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 09-Jul-2004  
C:Accession: A27041  
R:Haller, J.; Cote, S.; Broenner, G.; Jaecckle, H.  
Genes Dev. 1, 862-867, 1987  
A:Title: Dorsal and neural expression of a tyrosine kinase-related Drosophila gene during embryogenesis  
A:Reference number: A27041; MUID:88112827; PMID:3428600  
A:Accession: A27041  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-753 <HAL>  
A:Cross-references: UNIPROT:P14083; UNIPARC:UPI0000136FEA  
C:Genetics:  
A:Gene: dtkr  
A:Cross-references: FlyBase:FBgn0003715  
A:Map position: 2R, 60F1  
A:Introns: 453/1; 497/1  
C:Keywords: autophosphorylation; glycoprotein; phosphoprotein (covalent) #status predicted  
F:9,65,187,223,224,250,611,660/Binding site: carbohydrate (Asn) (by autophosphorylation) #status predicted  
F:744/Binding site: phosphate (Tyr) (covalent)

Query Match 64.8%; Score 35; DB 2; Length 753;  
Best Local Similarity 61.5%; Pred. No. 2.8e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13  
|: |||||  
Db 293 AQMHAAAAAAVAA 305

## RESULT 33

S35681  
ESG protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 05-Oct-2004  
C:Accession: S35681; S34162  
R:Miyaoka, H.; Choudhury, B.K.; Hou, E.W.; Li, S.S.L.  
Eur. J. Biochem. 216, 343-352, 1993  
A:Title: Molecular cloning and expression of mouse and human cDNA encoding AES and ESG proteins  
A:Reference number: S35678; MUID:93373944; PMID:8365415  
A:Accession: S35681  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-771 <MIY>  
A:Cross-references: UNIPROT:Q08122; UNIPARC:UPI0000027DAA; EMBL:X73360; NID:g313235; PID:g313235; PIDN:CAA66557.1  
C:Keywords: phosphoprotein  
F:527-560/Domain: WD repeat homology <WD1>  
F:613-646/Domain: WD repeat homology <WD3>  
F:695-728/Domain: WD repeat homology <WD4>  
F:736-769/Domain: WD repeat homology <WD5>

Query Match 64.8%; Score 35; DB 2; Length 771;  
Best Local Similarity 61.5%; Pred. No. 2.9e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
QY      1 ASMSAASASMAA 13
      : |||||: ||
Db      396 SMSAASASMAA 408

RESULT 34
G64887
probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: G64887; T09189
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64887
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1122 <BLAT>
A:Cross-references: UNIPROT:P76072; UNIPARC:UPI0000168098; GB:AE000234; GB:U00096; NID:9278503
A:Experimental source: strain K-12, substrain MG1655
R:Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Saito, moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A:Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A:Reference number: Z16603; MUID:97251357; PMID:9097039
A:Accession: T09189
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 3-1122 <AJB>
A:Cross-references: UNIPARC:UPI00001360ED; GB:AE000234; GB:U00096; NID:gl787633; PIDN:AF3008

Query Match      64.8%; Score 35; DB 2; Length 1122;
Best Local Similarity 75.0%; Pred. No. 4.1e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 ASMSAASASMAA 12
      : |||||: ||
Db      255 ASSSASASMAA 266

RESULT 35
AF3008
polyketide synthase Atu3672 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AF3008
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, J.E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF3008
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1520 <KUR>
A:Cross-references: UNIPROT:Q8U9Q4; UNIPARC:UPI00000D22BA; GB:AE008689; PIDN:AAL44484.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3672
A:Map position: linear chromosome

Query Match      64.8%; Score 35; DB 2; Length 1520;
Best Local Similarity 72.7%; Pred. No. 5.3e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 MSASASASMAA 13
      : |||||: ||
Db      484 LSARSASASMAA 494

RESULT 36
G98275
hypothetical protein AGR_L_2329 [imported] - Agrobacterium tumefaciens (strain C58, Ceratophyllum demersum)
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: G98275
R:Goodner, B.; Hinkley, C.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens strain C58.
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G98275
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1520 <KUR>
A:Cross-references: UNIPROT:Q8U9Q4; UNIPARC:UPI00000D22BA; GB:AE007870; PIDN:AAK99729.1;
C:Genetics:
A:Gene: AGR_L_2329
A:Map position: linear chromosome

Query Match      64.8%; Score 35; DB 2; Length 1520;
Best Local Similarity 72.7%; Pred. No. 5.3e+02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      3 MSASASASMAA 13
      : |||||: ||
Db      484 LSARSASASMAA 494

RESULT 37
T33152
hypothetical protein T04D1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T33152
R:Davidson, S.; Wohldmann, P.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of C. elegans cosmid T04D1.
A:Reference number: Z21292
A:Accession: T33152
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2957 <DAV>
A:Cross-references: UNIPROT:061845; UNIPARC:UPI000007A573; EMBL:AF067617; PIDN:AAC17555
A:Experimental source: strain Bristol N2; clone T04D1
C:Genetics:
A:Gene: CESP:T04D1.4
A:Map position: 1
A:Introns: 122/3; 293/3; 515/3; 1205/2; 1577/1; 2221/3; 2776/1; 2864/3

Query Match      64.8%; Score 35; DB 2; Length 2957;
Best Local Similarity 61.5%; Pred. No. 9.8e+02;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 ASMSAASASMAA 13
      : |||||: ||
Db      2844 ASMAAASASMAA 2856

RESULT 38
SXAD12
hexon-associated protein (IX) - human adenovirus 12
C:Species: Mastadenovirus h12 (human adenovirus 12)
C>Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 09-Jul-2004
C:Accession: A90814; A93745; S33931; A03855
R:Bos, J.L.; Polder, L.J.; Bernards, R.; Schrier, P.I.; van den Elsen, P.J.; van der Eb A.H.
Cell 27, 121-131, 1981
A:Title: The 2.2 kb E1b mRNA of human Ad12 and Ad5 codes for two tumor antigens starting at different reading frames.
A:Reference number: A90814; MUID:82115327; PMID:7326748
```

A;Accession: A90814  
A:Molecule type: mRNA  
A;Residues: 1-144 <BOS>  
A;Cross-references: UNIPROT:P03284; UNIPARC:UPI000004007E  
R;Kimura, T.; Sawada, Y.; Shinawawa, M.; Shimizu, Y.; Shiroki, K.; Shimojo, H.; Sugisaki  
Nucleic Acids Res. 9, 6571-6589, 1981  
A;Title: Nucleotide sequence of the transforming early region Elb of adenovirus type 12  
A;Reference number: A93745; MUID:82105565; PMID:6275367  
A;Accession: A93745  
A:Molecule type: DNA  
A;Residues: 1-144 <KIM>  
A;Cross-references: UNIPARC:UPI000004007E; GB:X73487; NID:g313361; PIDN:CAA51880.1; PID:  
R;Sprenkel, J.  
submitted to the EMBL Data Library, June 1993  
A;Reference number: S33928  
A;Accession: S33931  
A;Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-144 <SPR>  
A;Cross-references: UNIPARC:UPI000004007E; EMBL:X73487; NID:g313361; PIDN:CAA51880.1; PI  
C;Genetics:  
A;Map position: 9.4-10.7  
C;Superfamily: adenovirus hexon-associated protein (IX)

Query Match 63.0%; Score 34; DB 1; Length 144;  
Best Local Similarity 61.5%; Pred. No. 90;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
|||:|||||  
Db 71 ASAAASTARSMAA 83

RESULT 39

T03033  
probable cytochrome-c oxidase (EC 1.9.3.1) Vb chain precursor - rice mitochondrion  
C;Species: Mitochondrion Oryza sativa (rice)  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C;Accession: T03033  
R;Kadowaki, K.; Kubo, N.; Ozawa, K.; Hirai, A.  
EMBO J. 15, 6652-6661, 1996  
A;Title: Targeting presence acquisition after mitochondrial gene transfer to the nuclei  
A;Reference number: Z14829; MUID:97133294; PMID:8978691  
A;Accession: T03033  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A;Residues: 1-169 <KAD>  
A;Cross-references: UNIPROT:P92683; UNIPROT:Q8LQX9; UNIPARC:UPI000008B4C4; EMBL:D85381;  
A;Experimental source: cultivar Nipponbare  
C;Genetics:  
A;Gene: coxVb  
A;Genome: mitochondrion  
A;Introns: 55/1; 81/3; 121/1; 147/3  
C;Superfamily: mammalian cytochrome-c oxidase chain Vb  
C;Keywords: mitochondrion; oxidoreductase

Query Match 63.0%; Score 34; DB 2; Length 169;  
Best Local Similarity 69.2%; Pred. No. 1e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
|||:|||||  
Db 16 ASASASAAASGAA 28

RESULT 40

T46497  
hypothetical protein DKFZp434H247.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 09-Jul-2004  
C;Accession: T46497  
R;Koehler, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, January 2000

A;Reference number: Z23037  
A;Accession: T46497  
A;Status: preliminary  
A:Molecule type: mRNA  
A;Residues: 1-211 <AAA>  
A;Cross-references: UNIPROT:Q9NTF4; UNIPARC:UPI000006CFPC; EMBL:AL137304  
A;Experimental source: adult testis; clone DKFZp434H247  
C;Genetics:  
A;Note: DKFZp434H247.1

Query Match 63.0%; Score 34; DB 2; Length 211;  
Best Local Similarity 69.2%; Pred. No. 1.3e+02;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
|||:|||||  
Db 14 ASQQQASAAASAA 26

Search completed: September 9, 2006, 23:01:47  
Job time : 18.8101 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:38:21 ; Search time 119.139 Seconds  
(without alignments)  
100.934 Million cell updates/sec

Title: US-10-617-568-5

Perfect score: 54

Sequence: 1 ASMSAASAAASWAA 13

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : UniProt 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	79.6	272	2	Q6CVL8_KLULA
2	43	79.6	285	2	Q6K4T3_ORYSA
3	42	77.8	207	2	Q2N062_9STRA
4	42	77.8	328	2	Q927J6_LISIN
5	42	77.8	417	2	Q9VJX5_DROME
6	42	77.8	910	2	Q3DVC9_STRAG
7	42	77.8	970	2	Q8DY17_STRA5
8	42	77.8	1130	2	Q3DFZ9_STRAG
9	42	77.8	1258	2	Q3K052_STRA1
10	42	77.8	1310	2	Q8E473_STRA3
11	42	77.8	1326	2	Q3D424_STRAG
12	41	75.9	324	2	Q6C1D7_YARLI
13	41	75.9	805	2	Q4P0P5_USTMA
14	41	75.9	1293	2	Q7S1A6_NEUCR
15	40	74.1	210	2	Q6NCBS_RHOPA
16	40	74.1	510	1	FKH DROME
17	40	74.1	510	2	Q32KE3_DROME
18	40	74.1	1839	2	Q4B624_BURVI
19	39	72.2	181	2	Q3VRP3_ORYSA
20	39	72.2	201	2	Q342P2_RHOPA
21	39	72.2	216	2	Q6Z461_ORYSA
22	39	72.2	298	2	Q6CGK6_YARLI
23	39	72.2	301	2	Q41B28_GIBZE
24	39	72.2	378	2	Q62HU5_BURMA
25	39	72.2	378	2	Q63RN7_BURPS
26	39	72.2	395	2	Q8SNM6_ORYSA
27	39	72.2	441	2	Q45C13_9BURK
28	39	72.2	441	2	Q4LHN7_9BURK
29	39	72.2	470	2	Q9XUY1_CAEEL
30	39	72.2	629	2	Q2T1K2_BURTH
31	39	72.2	724	2	Q8NJ49_SACFI

## ALIGNMENTS

32	39	72.2	793	1	CAS DROME	07m3n8 drosophila
33	39	72.2	1014	2	Q2UGF0_ASPOR	Q2u6f0 aspergillus
34	39	72.2	4574	2	Q63LK9_BURPS	Q63lk9 burkholderi
35	38	70.4	129	2	Q2QLR1_ORYSA	Q2qlr1 oryza sativ
36	38	70.4	166	2	Q84JT2_ORYSA	Q84jt2 oryza sativ
37	38	70.4	169	2	Q2UKI7_ASPOR	Q2uky7 aspergillus
38	38	70.4	172	2	Q2XY88_DROYA	Q2xy88 drosophila
39	38	70.4	172	2	Q2XY89_DROYA	Q2xy89 drosophila
40	38	70.4	174	2	Q2XY87_DROER	Q2xy87 drosophila
41	38	70.4	328	2	Q4EFN1_LISMO	Q4efn1 listeria mo
42	38	70.4	328	2	Q4EP26_LISMO	Q4ep26 listeria mo
43	38	70.4	328	2	Q71WD5_LISMF	Q71wd5 listeria mo
44	38	70.4	328	2	Q8Y431_LISMO	Q8y431 listeria mo
45	38	70.4	330	2	Q89E13_BRAJA	Q89e13 bradyrhizob
46	38	70.4	332	2	Q2SYD1_BURTH	Q2syd1 burkholderi
47	38	70.4	343	2	Q61DL0_CAEER	Q61dl0 caenorhabdi
48	38	70.4	376	2	Q3BCG5_ARGAU	Q3bcg5 argiope aur
49	38	70.4	434	2	Q4B9B6_BURVI	Q4b9b6 burkholderi
50	38	70.4	472	2	Q4BIK8_BURVI	Q4bik8 burkholderi
51	38	70.4	508	1	CROC DROME	F32027 drosophila
52	38	70.4	508	2	Q53YH1_DROME	Q53yh1 drosophila
53	38	70.4	514	2	Q4PC76_USTMA	Q4pc76 ustilago ma
54	38	70.4	545	2	Q3BCG6_ARGAU	Q3bcg6 argiope aur
55	38	70.4	847	2	Q41EL9_GIBZE	Q41el9 gibberella
56	38	70.4	941	1	LOXC2_ORYSA	Q84yk8 oryza sativ
57	38	70.4	1090	2	Q4Q966_LEIMA	Q4q966 leishmania
58	38	70.4	3360	2	Q88XB6_LACPL	Q88xb6 lactobacill
59	37	68.5	76	2	Q69NX3_ORYSA	Q69nx3 oryza sativ
60	37	68.5	104	2	Q93T63_CORJE	Q93t63 corynebacte
61	37	68.5	131	2	Q4G1M4_ARGAU	Q4g1m4 argiope aur
62	37	68.5	151	2	Q5F3N8_AZOSE	Q5p3n8 azoarcus sp
63	37	68.5	197	2	Q5A210_CANAL	Q5a210 candida alb
64	37	68.5	199	2	Q944V7_PHYIN	Q944v7 phytophthor
65	37	68.5	234	2	Q3W4J1_9ACTO	Q3w4j1 frankia sp.
66	37	68.5	266	2	Q3J3S5_RHOSA	Q3j3s5 rhodobacter
67	37	68.5	285	2	Q2NGH6_9EURY	Q2ng6 methanosphe
68	37	68.5	286	2	Q84FA4_MYXXA	Q84fa4 myxococcus
69	37	68.5	290	2	Q62BA7_BURMA	Q62ba7 burkholderi
70	37	68.5	293	2	Q7XBR2_ORYSA	Q7xb2 oryza sativ
71	37	68.5	293	2	Q9FRM4_ORYSA	Q9frm4 oryza sativ
72	37	68.5	294	2	Q3JHA5_BURP1	Q3jha5 burkholderi
73	37	68.5	360	2	Q6L4G2_ORYSA	Q6l4g2 oryza sativ
74	37	68.5	375	2	Q664F0_YERP5	Q664f0 yersinia ps
75	37	68.5	375	2	Q8ZA42_YERPE	Q8za42 yersinia ps
76	37	68.5	376	2	Q8AUD4_ORYSA	Q8aud4 oryza sativ
77	37	68.5	394	2	Q4EXP5_LEIMA	Q4exp5 leishmania
78	37	68.5	426	2	Q7YU22_DROME	Q7yu22 drosophila
79	37	68.5	457	2	Q4NMG6_9MICC	Q4nm6 arthroabacte
80	37	68.5	462	2	Q98T82_CHICK	Q98t82 gallus gall
81	37	68.5	463	2	Q6UJ17_9CAUD	Q6uj17 burkholderi
82	37	68.5	518	2	Q62N55_BURMA	Q62n55 burkholderi
83	37	68.5	533	1	DNAA_BURPS	Q63yw5 burkholderi
84	37	68.5	533	2	Q3JXI6_BURP1	Q3jxi6 burkholderi
85	37	68.5	536	2	Q2STL6_BURTH	Q2stl6 burkholderi
86	37	68.5	540	2	Q47BW8_DECAR	Q47bw8 dechloromon
87	37	68.5	551	2	Q4NF37_9MICC	Q4nf37 arthroabacte
88	37	68.5	554	2	Q89FR7_BRAJA	Q89fr7 bradyrhizob
89	37	68.5	575	2	Q6U4B7_NILLU	Q6u4b7 nilaparvata
90	37	68.5	640	2	Q9VTW5_DROME	Q9vtw5 drosophila
91	37	68.5	653	2	Q2SUM3_BURTH	Q2sum3 burkholderi
92	37	68.5	735	2	Q6MY65_ASPFU	Q6my65 aspergillus
93	37	68.5	761	2	Q4WS26_ASPFU	Q4ws26 aspergillus
94	37	68.5	818	2	Q2SWB0_BURTH	Q2swb0 burkholderi
95	37	68.5	868	2	Q5SPK0_CRYNE	Q5spk0 cryptococcu
96	37	68.5	868	2	Q3KDV6_CRYNE	Q3kdv6 cryptococcu
97	37	68.5	1144	2	Q30PT0_THIDN	Q30pt0 thiomicrosp
98	37	68.5	2178	2	Q9KWR3_STRGN	Q9kwr3 streptococc
99	37	68.5	2601	2	Q4QGR7_LEIMA	Q4qgr7 leishmania
100	36	66.7	110	1	CBP1_ATECA	Q42740 ajellomyces

2

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN-CLIP 11262 / Serovar 6a;  
RX MEDLINE=21537279; PubMed=11679669; DOI=10.1126/science.1063447;  
RA Glaser P., Frangeul P., Buchrieser H., Rusniok C., Amend A.,  
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Delhou P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
RA Gautier L., Goebel W., Gomez-Jopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Maertens U., Kreft J., Kuhn M., Kunat F., Kurapkut G.,  
RA Madueno E., Maoutouram A., Mata Vicente J., Ng E., Nedjari H.,  
RA Madsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
RA Ruzman-Boland J.-A., Voss H., Weiland J., Cossart P.;  
RT "Comparative genomics of *Listeria species*.";  
RL Science 294:849-852(2001).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AL596173; CAC98018.1; -; Genomic\_DNA.  
DR PIR; AB1781; AB1781.  
DR ListiList; LIN2792; -.  
DR BioCyc; LINN1642; LIN2792-MONOMER; -.  
KW Complete proteome.  
SQ SEQUENCE 328 AA; 36670 MW; F2E1A0F2C6D98622 CRC64;  
Query Match 77.8%; Score 42; DB 2; Length 328;  
Best Local Similarity 76.9%; Pred. No. 98;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 ASMSAASAAASMAA 13  
|:||||:|||||  
Db 313 AAMSAAAAASAAA 325  
  
RESULT 5  
ID Q9VJX5 DROME PRELIMINARY; PRT; 417 AA.  
AC Q9VJX5;  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 01-MAY-2000, sequence version 1.  
DT 07-FEB-2006, entry version 19.  
DE CGL6850-PA.  
GN ORFNames=CGL6850, Dmel CGL6850;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferrieria S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Nusskern D.R., Pacleb J.M.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Spaden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RG Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AE003640; AAF53310.1; -; Genomic\_DNA.  
DR FlyBase; Fgn0032527; CGL6850.  
SQ SEQUENCE 417 AA; 44286 MW; 42EA717B5604565F CRC64;  
Query Match 77.8%; Score 42; DB 2; Length 417;

```

Best Local Similarity 76.9%; Pred. No. 1.2e+02;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13
DB 225 ASTSASASMAA 237
|||||:|||||
|||||:|||||

RESULT 6
Q3DFV9_STRAG PRELIMINARY; PRT; 910 AA.
AC Q3DFV9;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 1.
DE Cell wall surface anchor family protein.
GN ORFNames=SAJ_1458;
OS Streptococcus agalactiae 18RS21.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342613;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=18RS21;
RA Tettelin H., Masignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit Y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RA "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL: AAJ01000001; EAO63494.1; -: Genomic_DNA.
DR GO: GO:0009986; C:cell surface; IEA.
DR GO: GO:0005618; C:cell wall; IEA.
DR GO: GO:0009986; C:cell surface; IEA.
DR GO: GO:0005618; C:cell wall; IEA.
KW Cell wall; Peptidoglycan-anchor.
SQ SEQUENCE 910 AA; 94148 MW; C7F57665B2C99CFC CRC64;

Query Match 77.8%; Score 42; DB 2; Length 910;
Best Local Similarity 69.2%; Pred. No. 2.7e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13
DB 821 ASMSASTASMSA 833
|||||:|||||
|||||:|||||

RESULT 7
Q8DYL7_STRAS PRELIMINARY; PRT; 970 AA.
AC Q8DYL7_STRAS5
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, sequence version 18.
DE Cell wall surface anchor family protein.
GN OrderedLocusNames=SAG1462;
OS Streptococcus agalactiae serotype V.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;

```

```

[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=2603 V/R / ATCC BAA-611 / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit Y Ros I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RA "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL: AE014259; AAN00330.1; -: Genomic_DNA.
DR TIGR; SAG1462; -.
DR GO: GO:0009986; C:cell surface; IEA.
DR GO: GO:0005618; C:cell wall; IEA.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
SQ Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 970 AA; 99433 MW; 0E817044F6EE9CE6 CRC64;

Query Match 77.8%; Score 42; DB 2; Length 970;
Best Local Similarity 69.2%; Pred. No. 2.9e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13
DB 881 ASMSASTASMSA 893
|||||:|||||
|||||:|||||

RESULT 8
Q3DFZ9_STRAG PRELIMINARY; PRT; 1130 AA.
ID Q3DFZ9_STRAG
AC Q3DFZ9;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 5.
DE Cell wall surface anchor family protein.
GN Name=clfB; ORFNames=SAM_1428;
OS Streptococcus agalactiae CJB111.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342617;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CJB111;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Masignani V., Cieslewicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit Y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz M.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RA "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

```



```
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAJ001000008; EAO73831.1; -; Genomic_DNA.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
KW Cell wall; Peptidoglycan-anchor.
SQ SEQUENCE 1130 AA; 113624 MW; 264E398D72562BEE CRC64;

Query Match          77.8%; Score 42; DB 2; Length 1130;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13
| | | | | : | | | | |
Db 945 ASMSASTSASMSA 957

RESULT 9
ID Q3K052_STR1A1 PRELIMINARY; PRT; 1258 AA.
AC Q3K052;
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 08-NOV-2005, sequence version 1.
DE Cell wall surface anchor family protein.
DE OrderedLocusNames=SAK 1493; ORFNames=SAK_1493;
OS Streptococcus agalactiae serotype Ia.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=355315;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=A909 / ATCC 27591 / Serotype Ia;
RA PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Masiugnani V., Cieleswicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz J.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins C.E., Grandi K.J.,
RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,
RA Fraser C.M.;
RA "Genome analysis of multiple pathogenic isolates of Streptococcus
RT agalactiae: implications for the microbial 'pan-genome'.";
RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000114; AB454159.1; -; Genomic_DNA.
DR TIGR; SAK 1493; -
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 1258 AA; 124927 MW; AF0CC314FFB9B9EF CRC64;

Query Match          77.8%; Score 42; DB 2; Length 1258;
Best Local Similarity 69.2%; Pred. No. 3.8e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13
| | | | | : | | | | |
Db 849 ASMSASTSASMSA 861

-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAJ001000008; EAO73831.1; -; Genomic_DNA.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
KW Cell wall; Peptidoglycan-anchor.
SQ SEQUENCE 1130 AA; 113624 MW; 264E398D72562BEE CRC64;

Query Match          77.8%; Score 42; DB 2; Length 1130;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13
| | | | | : | | | | |
Db 945 ASMSASTSASMSA 957

RESULT 10
ID Q8E473_STR3A3 PRELIMINARY; PRT; 1310 AA.
AC Q8E473;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE Hypothetical protein gbs1529.
GN OrderedLocusNames=gbs1529;
OS Streptococcus agalactiae serotype III.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Meadek T., Zouine M., Couve E., Lalioui L., Foyart C., Trieu-Cuot P.,
RA Kunst F.;
RA "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AL766851; CAD47188.1; -; Genomic_DNA.
DR Sagalish; gbs1529; -
DR BioCyc; SAGA21110:GBS1529-MONOMER; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Complete proteome; Peptidoglycan-anchor.
SQ SEQUENCE 1310 AA; 129398 MW; 9FB7A8BE89B8F155 CRC64;

Query Match          77.8%; Score 42; DB 2; Length 1310;
Best Local Similarity 69.2%; Pred. No. 3.9e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13
| | | | | : | | | | |
Db 781 ASMSASTSASMSA 793

-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAJ001000008; EAO73831.1; -; Genomic_DNA.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
KW Cell wall; Peptidoglycan-anchor.
SQ SEQUENCE 1130 AA; 113624 MW; 264E398D72562BEE CRC64;

Query Match          77.8%; Score 42; DB 2; Length 1130;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13
| | | | | : | | | | |
Db 945 ASMSASTSASMSA 957

RESULT 11
ID Q3D424_STRAG PRELIMINARY; PRT; 1326 AA.
AC Q3D424;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Cell wall surface anchor family protein.
DE Name=sdrE; ORFNames=SAI_1566;
GN Streptococcus agalactiae H368.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=342615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H368;
RX PubMed=16172379; DOI=10.1073/pnas.0506758102;
RA Tettelin H., Masiugnani V., Cieleswicz M.J., Donati C., Medini D.,
RA Ward N.L., Angiuoli S.V., Crabtree J., Jones A.L., Durkin A.S.,
RA DeBoy R.T., Davidson T.M., Mora M., Scarselli M., Margarit y Ros I.,
RA Peterson J.D., Hauser C.R., Sundaram J.P., Nelson W.C., Madupu R.,
RA Brinkac L.M., Dodson R.J., Rosovitz J.J., Sullivan S.A.,
RA Daugherty S.C., Haft D.H., Selengut J., Gwinn M.L., Zhou L., Zafar N.,
RA Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K.J.,
RA Smith S., Utterback T.R., White O., Rubens C.E., Grandi G.,
```

RA Madoff L.C., Kasper D.L., Telford J.L., Wessels M.R., Rappuoli R.,  
 RA Fraser C.M.;  
 RT "genome analysis of multiple pathogenic isolates of *Streptococcus*  
 RL agalactiae: implications for the microbial 'pan-genome'.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 102:13950-13955(2005).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: AJ250100004; EAO78954.1; -; Genomic\_DNA.  
 DR GO: GO:0009986; C:cell surface; IEA.  
 DR GO: GO:0005618; C:cell wall; IEA.  
 KW Cell wall; Peptidoglycan-anchor.  
 SQ SEQUENCE 1326 AA; 131101 MW; 2866221DCFB528E CRC64;

Query Match 77.8%; Score 42; DB 2; Length 1326;  
 Best Local Similarity 69.2%; Pred. No. 4e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
 DB 685 ASMSAASASMAA 697  
 |||||:|||||:  
 |||||:|||||:

RESULT 12  
 ID Q6CID7\_YARLI PRELIMINARY; PRT; 324 AA.  
 AC Q6CID7;  
 DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.  
 DT 16-AUG-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 12.  
 DE Similar to tr|Q9C228 Neurospora crassa Related to CSI2 protein.  
 GN OrderedLocusNames=YALIOF17204g.  
 OS Yarrowia lipolytica (Candida lipolytica).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 OX NCBI\_TaxID=4952;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=CLIB 122 / E 150;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Tallia E.,  
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boissrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenn S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
 RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zénou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissensbach J.,  
 RA Winkler P., Souciet J.-L.;  
 RT "Genome evolution in yeasts."  
 RL Nature 430:35-44 (2004).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: CR382132; CAG78334.1; -; Genomic\_DNA.  
 KW Complete proteome.  
 SQ SEQUENCE 324 AA; 35095 MW; 69775D1A260C9908 CRC64;

Query Match 75.9%; Score 41; DB 2; Length 324;  
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
 Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 SMSAASASMAA 13

Db 179 SMSQAASMAA 190  

RESULT 13  
 Q4POP5\_USTWA  
 ID Q4POP5 PRELIMINARY; PRT; 805 AA.  
 AC Q4POP5;  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE Hypothetical protein.  
 GN ORFNames=UM06318.1;  
 OS Ustilago maydis 521.  
 OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;  
 OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.  
 OX NCBI\_TaxID=237631;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=521;  
 RA Birren B.W., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,  
 RA Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,  
 RA Arachchi H.M., Armbruster J., Bachantsang P., Baldwin J., Barry A.,  
 RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,  
 RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,  
 RA Calvo S.E., Camarata J., Campo K., Chang J., Cheshatsang Y.,  
 RA Citroen M., Collymore A., Considine T., Cook A., Cooke P., Corum B.,  
 RA Cuomo C., David R., Dawoe T., Degray S., Dodge S., Dooley K.,  
 RA Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,  
 RA Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,  
 RA Fitzgerald M., Foley K., Gage D., Galagan J.E., Gearin G., Gerre S.,  
 RA Gnirke A., Goyette A., Graham J., Grandbois E., Gvaltsen K., Hafez N.,  
 RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,  
 RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,  
 RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseis M., Karlsson E.,  
 RA Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,  
 RA Lana D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,  
 RA Lindblad-Toh K., Liu X., Lokytsang T., Lokytsang Y., Lucien O.,  
 RA Lui A., Ma L.-J., Mabbitt R., MacDonald J., MacLean C., Major J.,  
 RA Manning J., Marabellia R., Maru K., Matthews C., Mauceli E.,  
 RA McCarthy M., McDonough S., McGhee T., Meldrum J., Menes L.,  
 RA Mesirov J., Mihailev A., Mihova T., Mikkelsen T., Menga V., Moru K.,  
 RA Mores J., Mulrain L., Munson G., Naylor J., Neves C., Nguyen C.,  
 RA Nguyen N., Nguyen T., Nicol R., Nielsen C.B., Nizzari M., Norbu C.,  
 RA Norbu N., O'Donnell P., Okoawo O., O'Leary S., Omotosho B.,  
 RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pigani B.,  
 RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,  
 RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,  
 RA Rutnan M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,  
 RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,  
 RA Spencer B., Stalker J., Stange-Thomann N., Stavropoulos S.,  
 RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,  
 RA Tenzing P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,  
 RA Towey S., Tsanla T., Tsomo N., Vallee D., Vassiliev H.,  
 RA Venkataraman V.S., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,  
 RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,  
 RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembek L.,  
 RA Zimmer A., Zody M., Lander E.S.;  
 RT "The genome sequence of *Ustilago maydis*."  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: AAC01000242; EAK82947.1; -; Genomic DNA.  
 DR GO: GO:0044249; P:cellular biosynthesis; IEA.  
 DR InterPro: IPR000649; IF-2B related.  
 DR PANTHER: PTHR10233; IF-2B; 1.  
 DR Pfam: PF01008; IF-2B; 1.  
 KW Hypothetical protein.

```
SQ SEQUENCE 805 AA; 86248 MW; 5A4CC28D1060ACD6 CRC64;

Query Match 75.9%; Score 41; DB 2; Length 805;
Best Local Similarity 76.9%; Pred. No. 3.5e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
DB 131 ASASAAASMAA 143

RESULT 14
Q7S1A6_NEUCR PRELIMINARY; PRT; 1293 AA.
ID Q7S1A6;
AC Q7S1A6;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Predicted protein.
GN ORFNames=NCU07514.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
STRAIN=74-OR23-1A / FGSC 987;
RC MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A.,
RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
RA Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,
RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamysseelis M.,
RA Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S.,
RA Rasmussen G., Metznerberg R.L., Perkins D.D., Kroken S., Cogoni C.,
RA Macino G., Catcheside D.E.A., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
RA Yarden O., Piamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.W.;
RT "The genome sequence of the filamentous fungus Neurospora crassa.";
RL Nature 422:859-868(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBB whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AABX01000455; EAA29126.1; -; Genomic_DNA.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN 1.
SQ SEQUENCE 1293 AA; 136804 MW; C85AF94F062EB449 CRC64;

Query Match 75.9%; Score 41; DB 2; Length 1293;
Best Local Similarity 84.6%; Pred. No. 5.6e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
DB 1067 ASASAAASMAA 1079

RESULT 15
Q6NCES_RHOPA PRELIMINARY; PRT; 210 AA.
ID Q6NCES5;
AC Q6NCES5;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Predicted protein.
GN ORFNames=NCU07514.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
STRAIN=74-OR23-1A / FGSC 987;
RC MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A.,
RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
RA Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,
RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamysseelis M.,
RA Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S.,
RA Rasmussen G., Metznerberg R.L., Perkins D.D., Kroken S., Cogoni C.,
RA Macino G., Catcheside D.E.A., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
RA Yarden O., Piamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.W.;
RT "The genome sequence of the filamentous fungus Neurospora crassa.";
RL Nature 422:859-868(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBB whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AABX01000455; EAA29126.1; -; Genomic_DNA.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN 1.
SQ SEQUENCE 1293 AA; 136804 MW; C85AF94F062EB449 CRC64;

Query Match 75.9%; Score 41; DB 2; Length 1293;
Best Local Similarity 84.6%; Pred. No. 5.6e+02;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
DB 1067 ASASAAASMAA 1079

RESULT 15
Q6NCES5_RHOPA PRELIMINARY; PRT; 210 AA.
ID Q6NCES5;
AC Q6NCES5;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Predicted protein.
GN ORFNames=NCU07514.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]

NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
STRAIN=74-OR23-1A / FGSC 987;
RC MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Bell-Pedersen D., Nelson M.A.,
RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,
RA Zelter A., Schulte U., Kothe G.O., Jedd G., Mewes H.-W., Staben C.,
RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,
RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamysseelis M.,
RA Mauceli E., Bielke C., Rudd S., Frishman D., Krystofova S.,
RA Rasmussen G., Metznerberg R.L., Perkins D.D., Kroken S., Cogoni C.,
RA Macino G., Catcheside D.E.A., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,
RA Yarden O., Piamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.W.;
RT "The genome sequence of the filamentous fungus Neurospora crassa.";
RL Nature 422:859-868(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBB whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AABX01000455; EAA29126.1; -; Genomic_DNA.
DR InterPro; IPR000886; ER target S.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN 1.
SQ SEQUENCE 1293 AA; 136804 MW; C85AF94F062EB449 CRC64;

Query Match 74.1%; Score 40; DB 2; Length 210;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMA 12
DB 173 ASASAAASAAASIA 184

RESULT 16
FKH DROME STANDARD; PRT; 510 AA.
ID FKX DROME;
AC P14734; Q9VAV0;
DT 01-APR-1990, integrated into UniProtKB/Swiss-Prot.
DT 01-APR-1990, sequence version 1.
DT 07-FEB-2006, entry version 57.
DE Fork head protein.
GN Name=fkh; ORFNames=CGI10002;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC MEDLINE=89249328; PubMed=2566386; DOI=10.1016/0092-8674(89)90133-5;
RA Weigel D., Juergens G., Kuetner F., Seifert E., Jacek H.;
RT "The homeotic gene fork head encodes a nuclear protein and is
RT expressed in the terminal regions of the Drosophila embryo.";
RL Cell 57:645-658(1989).
RN [2]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Berkley;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
```

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Nusskern D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.M., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler M.P., Shen H.,  
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Yao Q.A.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195 (2000).  
RN [3]  
RP GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettecourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.22 (2002).  
CC -!- FUNCTION: Fkh promotes terminal as opposed to segmental  
CC development. In the absence of fkh, this developmental switch does  
CC not occur. The nuclear localization of the fkh protein suggest  
CC that fkh regulates the transcription of other, subordinate, genes.  
CC -!- INTERACTION:  
CC Q9V860:CG6459; NbExp=1; IntAct=EBI-124027, EBI-151216;  
CC -!- SUBCELLULAR LOCATION: Nucleus.  
CC -!- SIMILARITY: Contains 1 fork-head DNA-binding domain.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: J03177; AAA28535.1; -; Genomic DNA.  
DR EMBL; AE003766; AAF56798.1; -; Genomic DNA.  
DR PIR; A32380; A32380.  
DR HSP; Q99958; I05V.  
DR SRF; P14734; 210-300.  
DR IntAct; P14734; -.  
DR TRANSFAC; T01054; -.  
DR Ensembl; CG10002; *Drosophila melanogaster*.  
DR Flybase; FBgn0000659; fkh.  
DR BioCyc; DMEL-XXX-02:DMEL-XXX-02-01382-MONOMER; -.  
DR GO; GO:0005515; P:protein binding; IPI.  
DR GO; GO:0006916; P:anti-apoptosis; IMP.  
DR GO; GO:0007439; P:ectodermal gut development; TAS.  
DR GO; GO:0001706; P:endoderm formation; TAS.  
DR GO; GO:0007443; P:Malpighian tubule morphogenesis; TAS.  
DR GO; GO:0000122; P:negative regulation of transcription from R. .; TAS.  
DR GO; GO:0007435; P:salivary gland morphogenesis; IMP.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR InterPro; IPR011991; Wing\_hlx\_DNA\_bd.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.

DR PROSITE; PS00657; FORK HEAD 1; 1.  
DR PROSITE; PS00658; FORK HEAD 2; 1.  
DR PROSITE; PS00039; FORK HEAD 3; 1.  
KW Complete proteome; Developmental protein; DNA-binding;  
KW Nuclear protein; Transcription; Transcription regulation.  
FT CHAIN 1 510 Fork head protein.  
FT FTID=PRO\_0000091896.  
FT DNA BIND 209 300 Fork-head.  
FT COMPIAS 21 40 Gly-rich.  
FT COMPIAS 340 356 His-rich.  
SQ SEQUENCE 510 AA; 54286 MW; 2B86E807CBF7881B CRC64;  
Query Match 74.1%; Score 40; DB 1; Length 510;  
Best Local Similarity 69.2%; Pred. No. 3.2e+02;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ASMSAASASMAA 13  
Db 101 ASMSASMSASMSA 113  
RESULT 17  
Q32KE3 DROME PRELIMINARY; PRT; 510 AA.  
ID Q32KE3 DROME PRELIMINARY; PRT; 510 AA.  
AC Q32KE3.  
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.  
DT 06-DEC-2005, sequence version 1.  
DT 07-MAR-2006, entry version 4.  
DE RE06859p.  
GN Name=fkh.  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1] NUCLEOTIDE SEQUENCE.  
RP STRAIN=Berkley.  
RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleb J.,  
RA Park S., Wan K., Yu C., Celniker S.;  
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nucleus (By similarity).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; BT023936; ABB36440.1; -; mRNA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; P:transcription factor activity; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
KW DNA-binding; Nuclear protein.  
SQ SEQUENCE 510 AA; 54242 MW; 3089574C2BE7981B CRC64;  
Query Match 74.1%; Score 40; DB 2; Length 510;  
Best Local Similarity 69.2%; Pred. No. 3.2e+02;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ASMSAASASMAA 13  
Db 101 ASMSASMSASMSA 113  
RESULT 18  
Q4B6Z4 BURVI PRELIMINARY; PRT; 1839 AA.  
ID Q4B6Z4 BURVI PRELIMINARY; PRT; 1839 AA.  
AC Q4B6Z4.  
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.  
DT 13-SEP-2005, sequence version 1.  
DT 07-FEB-2006, entry version 2.  
DE Hypothetical protein.  
GN ORFNames=Bcep1808DRAFT\_7796;  
OS Burkholderia vietnamiensis G4.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;

```

OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
OX NCBI_TaxID=269482;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=G4;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Burkholderia
RT vietnamiensis G4.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=G4;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia vietnamiensis
RT G4.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=G4;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAEH02000103; EAM25629.1; -; Genomic_DNA.
DR InterPro; IPR003006; IG_MHC.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 1839 AA; 185155 MW; 0DADC47988CEC299 CRC64;
Query Match 74.1%; Score 40; DB 2; Length 1839;
Best Local Similarity 76.9%; Pred. No. 1.2e+03;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASMSAASAASMAA 13
: |||||
DB 1750 AASSAASASQA 1762

RESULT 19
QSVRP3_ORYSA PRELIMINARY; PRT; 181 AA.
AC QSVRP3_
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DT 21-FEB-2006, entry version 7.
DE Hypothetical protein P0679C08.7-2 (Hypothetical protein P0493C11.25-
DE 2)
GN Name=P0679C08.7-2; Synonyms=P0493C11.25-2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Bep clade;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
RT clone.P0679C08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC

Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
NCBI_TaxID=269482;
[1]
NUCLEOTIDE SEQUENCE.
STRAIN=G4;
US DOE Joint Genome Institute (JGI-PGF);
Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
Hamon N., Israni S., Pittluck S., Richardson P.;
Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
-!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
-----
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AAEH02000103; EAM25629.1; -; Genomic_DNA.
InterPro; IPR003006; IG_MHC.
PROSITE; PS00290; IG_MHC; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 1839 AA; 185155 MW; 0DADC47988CEC299 CRC64;
Query Match 74.1%; Score 40; DB 2; Length 1839;
Best Local Similarity 76.9%; Pred. No. 1.2e+03;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASMSAASAASMAA 13
: |||||
DB 1750 AASSAASASQA 1762

RESULT 19
QSVRP3_ORYSA PRELIMINARY; PRT; 181 AA.
AC QSVRP3_
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DT 21-FEB-2006, entry version 7.
DE Hypothetical protein P0679C08.7-2 (Hypothetical protein P0493C11.25-
DE 2)
GN Name=P0679C08.7-2; Synonyms=P0493C11.25-2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Bep clade;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
RT clone.P0679C08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC

```

```

RT clone:P0493C11.";
Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP002542; BAD67872.1; -; Genomic_DNA.
DR EMBL; AP000559; BAD67675.1; -; Genomic_DNA.
DR Gramene; QSVRP3; -.
KW Hypothetical protein.
SQ SEQUENCE 181 AA; 18397 MW; A5CAB7C54081C7D0 CRC64;
Query Match 72.2%; Score 39; DB 2; Length 181;
Best Local Similarity 69.2%; Pred. No. 1.7e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASMSAASAASMAA 13
: |||||
DB 50 AAMSASASAAAA 62

RESULT 20
Q342F2_RHOPA PRELIMINARY; PRT; 201 AA.
ID Q342F2_RHOPA
AC Q342F2;
DT 06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 06-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Cell wall surface anchor family protein (Fragment).
DE ORFNames=RPCDRAFT_0879;
OS Rhodopseudomonas palustris Bis18.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=316056;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Bis18;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome and assembly of Rhodopseudomonas
RT palustris Bis18.";
RL Submitted (OCT-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Bis18;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Rhodopseudomonas palustris
RT Bis18.";
RL Submitted (NOV-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AALR01000022; EAP10010.1; -; Genomic_DNA.
FT NON TER 1
SQ SEQUENCE 201 AA; 17690 MW; A28DC1EFB43DF437 CRC64;
Query Match 72.2%; Score 39; DB 2; Length 201;
Best Local Similarity 76.9%; Pred. No. 1.8e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASMSAASAASMAA 13
: |||||
DB 146 ATRSASAASAAA 158

RESULT 21
Q62461_ORYSA

```

```

ID Q6Z461_ORYSA PRELIMINARY; PRT; 216 AA.
AC Q6Z461;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Putative squamosa-promoter binding protein.
GN Name=P0430F03.47;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=39947;
RN NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone:P0430F03.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AP005186; BAC84006.1; -; Genomic_DNA.
DR Gramine; Q62461; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR004333; SBP.
DR Pfam; PF03110; SBP; 1.
SQ SEQUENCE 216 AA; 22044 MW; AF74A6B47B498E60 CRC64;
Query Match 72.2%; Score 39; DB 2; Length 216;
Best Local Similarity 69.2%; Pred. No. 2e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ASMSAASAAASMAA 13
:|:|:|:|:|:|
Db 12 SSSSSSSAASMAA 24

RESULT 22
O6CGK6_YARLI PRELIMINARY; PRT; 298 AA.
AC O6CGK6;
DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Similarity.
GN OrderedLocusNames=YAL10A18524g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CLIB 122 / E 150;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Tallia E.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boissame A., Boyer J., Cathelico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Winkler P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
```

```

CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; CR382127; CAG841139.1; -; Genomic_DNA.
DR InterPro; IPR001778; POA allergen_C.
DR PRINTS; PR00833; POALLERGEN.
KW Complete proteome.
SQ SEQUENCE 298 AA; 29325 MW; 50660B6CAC94A883 CRC64;
Query Match 72.2%; Score 39; DB 2; Length 298;
Best Local Similarity 76.9%; Pred. No. 2.7e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 ASMSAASAAASMAA 13
:|:|:|:|:|:|
Db 222 AASSAASAAASMAA 234

RESULT 23
O4IBZ8_GIBZE PRELIMINARY; PRT; 301 AA.
AC O4IBZ8;
DT 16-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 16-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=FG05260.1;
OS Gibberella zeae (Fusarium graminearum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=5518;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=PH-1 / NRRL 31084;
RA Birren B.W., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D.,
RA Galagan J.E., Gardyna S., Gnerre S., Graham L., Grand-Pierre N.,
RA Hafez N., Hagopian D., Hagos B., Hall J., Horton L., Hulme W.,
RA Iliev I., Jaffe D., Johnson R., Jones C., Kamal M., Kamat A.,
RA Karatas A., Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G.,
RA Lui A., Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J.,
RA Manning J., Matthews C., Mauceli E., Murphy T., Naylor J., Nguyen C.,
RA Meneus L., Mihova T., Mieng V., Norbu C., O'Connor T., O'Donnell P.,
RA O'Neil D., Oliver J., Peterson K., Phunkhang P., Pierre N.,
RA Purcell S., Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C.,
RA Rogov P., Roman J., Schauer S., Schuback R., Seaman S., Severy P.,
RA Smirnov S., Smith C., Spencer B., Stange-Thomann N., Stojanovic N.,
RA Stubbs M., Tallamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.S.;
RT "Fusarium graminearum genome sequence.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AACM0100212; EAA75496.1; -; Genomic_DNA.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 301 AA; 32281 MW; E5838FC4C947BA3B CRC64;
Query Match 72.2%; Score 39; DB 2; Length 301;
Best Local Similarity 81.8%; Pred. No. 2.7e+02;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```

Oy 3 MSAASAASMAA 13
Db 46 LSAASAASMAA 56

RESULT 24
Q62HU5_BURMA
ID Q62HU5_BURMA PRELIMINARY; PRT; 378 AA.
AC Q62HU5_
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.
DT 25-OCT-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Hypothetical protein.
GN OrderedLocusNames=BMA2157;
OS Burkholderia mallei (Pseudomonas mallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=13373;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 23344;
RX PubMed=15377793; DOI=10.1073/pnas.0403306101;
RA Nierman W.C., DeShazer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T.V., Ulrich R.L., Ronning C.M., Brinkac L.M.,
RA Daugherty S.C., Davidsen T.D., DeBoy R.T., Dimitrov G., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Haft D.H., Khouri H.M., Kolonay J.F.,
RA Madupu R., Mohammed Y., Nelson W.C., Radune D., Romero C.M.,
RA Sarria S., Selengut J., Shamblin C., Sullivan S.A., White O., Yu Y.,
RA Zafar N., Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14246-14251(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP000010; AAU49932.1; -; Genomic_DNA.
DR TIGR; BMA2157; -;
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR004455; NADPoxred_F420.
DR Pfam; PF03807; F420 oxidored; I.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 378 AA; 38312 MW; E714B642803AD906 CRC64;

Query Match 72.2%; Score 39; DB 2; Length 378;
Best Local Similarity 69.2%; Pred. No. 3.4e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ASMSAASAASMAA 13
Db 36 ASRSSASAALAA 48

RESULT 26
Q9SNM6_ORYZA
ID Q9SNM6_ORYZA PRELIMINARY; PRT; 395 AA.
AC Q9SNM6; Q84Z21;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-MAY-2000, sequence version 1.
DT 21-FEB-2006, entry version 20.
DE Putative OSMAC7 protein (Putative NAM (No apical meristem) protein).
GN Name=P0493C11.25-1; Synonyms=134P10.2, P0679C08.7-1;
OR Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
clone:P0493C11.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
clone:P0679C08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Ma J., Sanmiguel P.J., Dubcovsky J., Shiloff B.A., Rostoks N.,
RA Jiang Z., Busso C.S., Kleinhofs A., Devos K.M., Ramakrishna W.,
RA Bennetzen J.L.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF000559; BAA84803.1; -; Genomic DNA.
DR EMBL; AF002542; BAB19365.1; -; Genomic DNA.
DR EMBL; AF488413; AAO33144.1; -; Genomic_DNA.
DR HSSP; Q9C932; 1U77.
DR Gramene; Q9SNM6; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0045449; P:regulation of transcription; IEA.
DR InterPro; IPR003441; NAM.

```

```

DR Pfam; PF02365; NAM; 1.
DR PROSITE; PSS1005; NAC; 1.
SQ SEQUENCE 395 AA; 42027 MW; 907916BE379EE14C CRC64;

Query Match 72.2%; Score 39; DB 2; Length 395;
Best Local Similarity 69.2%; Pred. No. 3.6e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
Db 264 AMSASASAAAAA 276
|:||||:||||: ||
|:||||:||||: ||

RESULT 27
Q45CI3_9BURK
ID Q45CI3_9BURK PRELIMINARY; PRT; 441 AA.
AC Q45CI3;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Hypothetical protein.
GN ORFNames=BcenDRAFT_5768;
OS Burkholderia cenocepacia AU 1054.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; Burkholderia cepacia complex.
NCBI_TaxID=331271;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AU 1054;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hammon N., Israni S., Pitluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Burkholderia cenocepacia
RT AU 1054.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
[2]
RN
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AU 1054;
RG US DOE Joint Genome Institute (JGI-ORNL);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Burkholderia cenocepacia
RT AU 1054.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AAHT01000002; EAM14373.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 441 AA; 46730 MW; 81765FDA8EF6B8D5 CRC64;

Query Match 72.2%; Score 39; DB 2; Length 441;
Best Local Similarity 76.9%; Pred. No. 4e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13
Db 139 SSASASAAASAAA 151
|:||||||| ||
|:||||||| ||

RESULT 28
Q4LHN7_9BURK
ID Q4LHN7_9BURK PRELIMINARY; PRT; 441 AA.
AC Q4LHN7;
DT 02-AUG-2005, integrated into UniProtKB/TrEMBL.
DT 02-AUG-2005, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Hypothetical protein.
GN ORFNames=Bcen2424DRAFT_0493;
OS Burkholderia cenocepacia HI2424.

```



```
DR InterPro; IPR0013604; Znf_U1.
DR InterPro; IPR013085; Znf_U1-C.
DR Pfam; PF06220; zf-U1; 1.
DR SMART; SM00451; Znf_U1; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 470 AA; 53845 MW; 91F3A5A0A433E37A CRC64;

Query Match 72.2%; Score 39; DB 2; Length 470;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMA 12
   |||:|||||
Db 66 AKMEAAAAASMA 77

RESULT 30
Q2TIK2_BURTH PRELIMINARY; PRT; 629 AA.
AC Q2TIK2;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DE 07-FEB-2006, entry version 2.
DE Gamma-glutamyltransferase.
GN ORFNames=BrH_10389;
OS Burkholderia thailandica E264.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=271848;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=E264;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Palmer N., Dodson R.,
RA Hickey E.K., Gwin M., Dougherty B., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van-Vugt R., Adams M.D., Gocayne J.D., Weidman J., Utterback T.,
RA Wathey L., McDonald L., Artiach P., Bowman C., Garland S., Fujii C.,
RA Cotton M.D., Horst K., Tomb J.-F., Roberts K., Hatch B., Smith H.O.,
RA Venter J.C.;
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CP0000086; ABC37984.1; -; Genomic DNA.
DR GO; GO:0016740; F:transferase activity; IEA.
KW Transferase.
SQ SEQUENCE 629 AA; 65145 MW; 305D2666A67EC6EE CRC64;

Query Match 72.2%; Score 39; DB 2; Length 629;
Best Local Similarity 76.9%; Pred. No. 5.7e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13
   |||:|||||
Db 42 ASASAAASAAAS 54

RESULT 31
Q8NJ49_SACFI PRELIMINARY; PRT; 724 AA.
AC Q8NJ49;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DE 07-FEB-2006, entry version 15.
DE Acetolactate synthase (EC 4.1.3.18).
GN Name=tlv2;
OS Saccharomycopsis fibuligera (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.
OX NCBI_TaxID=4944;
RN [1]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=8014;
RA Ng L.L.H., Nga B., Tan K.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ491151; CAD36014.1; -; Genomic DNA.
DR HSSP; P07342; INOH.
DR SMR; Q8NJ49; 117-724.
DR GO; GO:0003984; F:acetolactate synthase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:00030976; F:thiamin pyrophosphate binding; IEA.
DR GO; GO:0009082; P:branched chain family amino acid biosynthesis; IEA.
DR InterPro; IPR012846; Acetolac_syn_lg.
DR InterPro; IPR004407; Acetolac_syn_lg.
DR InterPro; IPR000399; TPP_bd_
DR InterPro; IPR012001; TPP_bd_enzyme_N.
DR InterPro; IPR011766; TPP_enzyme_bd_C.
DR InterPro; IPR012000; TPP_enzyme_M.
DR PANTHER; PTHR18968.SF61; Acolac_synthlrg; 1.
DR Pfam; PF02775; TPP_enzyme_C; 1.
DR Pfam; PF00205; TPP_enzyme_M; 1.
DR Pfam; PF02776; TPP_enzyme_N; 1.
DR PIRSF; PIRSF00108; Acetolac_syn_lg; 1.
DR PIRSF; PIRSF001370; Thdp_depend_ac1; 1.
DR TIGRFAMs; TIGR00118; acolac_lg; 1.
DR PROSITE; PS00187; TPP_ENZYMES; 1.
KW Lyase.
SQ SEQUENCE 724 AA; 77781 MW; 678EC6021FC874DF CRC64;

Query Match 72.2%; Score 39; DB 2; Length 724;
Best Local Similarity 76.9%; Pred. No. 6.6e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13
   |||:|||||
Db 87 ASASASSAASAAA 99

RESULT 32
CAS_DROME STANDARD; PRT; 793 AA.
AC Q7MJM8; Q9VNH0;
DT 24-MAY-2004, integrated into UniProtKB/Swiss-Prot.
DT 24-MAY-2004, sequence version 2.
DE Transcription factor castor (Protein ming).
DE Transcription factor castor (Protein ming).
GN Name=cast; Synonyms=ming; ORFNames=CG2102;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_TaxID=7227;
RP NUCLEOTIDE SEQUENCE, FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX PubMed=1418995; DOI=10.1016/0896-6273(92)90234-5;
RA Mellerick D.M., Kassisi J.A., Zhang S.-D., Odenwald W.F.;
RT "Castor encodes a novel zinc finger protein required for the
RT development of a subset of CNS neurons in Drosophila.";
RL Neuron 9:789-803(1992).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfammkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RA "The genome sequence of *Drosophila melanogaster*."  
RT Science 287:2185-2195(2000).  
RN [3]  
RP GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Battencourt B.R., Calniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.B., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.,  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review."  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RX STRAIN=Berkley; TISSUE=Embryo;  
RA MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
RA Rubin G.M., Calniker S.E.,  
RA Rubin G.M., Calniker S.E.,  
RT "A *Drosophila* full-length cDNA resource."  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE OF 281-660, FUNCTION, AND TISSUE SPECIFICITY.  
RC TISSUE=Embryo;  
RX PubMed=1339340;  
RA Cui X., Doe C.Q.,  
RT "ming is expressed in neuroblast sublineages and regulates gene  
RT expression in the *Drosophila* central nervous system."  
RL Development 116:943-952(1992).  
RN [6]  
RP TISSUE SPECIFICITY.  
RX PubMed=7588058;  
RA Cui X., Doe C.Q.,  
RT "The role of the cell cycle and cytokinesis in regulating neuroblast  
RT sublineage gene expression in the *Drosophila* CNS."  
RL Development 121:3233-3243(1995).  
RN [7]  
RP FUNCTION AS REPRESSOR, DNA-BINDING, AND SUBCELLULAR LOCATION.  
RX PubMed=9436984;

RA Kambadur R., Koizumi K., Stivers C., Nagle J., Poole S.J.,  
RA Odenwald W.F.,  
RT "Regulation of POU genes by castor and hunchback establishes layered  
RT compartments in the *Drosophila* CNS."  
RL Genes Dev. 12:246-260(1998).  
RN [8]  
RP TISSUE SPECIFICITY.  
RX PubMed=11335107; DOI=10.1016/S0925-4773(01)00312-4;  
RA Hitier R., Chaminade M., Preat T.,  
RT "The *Drosophila* castor gene is involved in postembryonic brain  
RT development."  
RL Mech. Dev. 103:3-11(2001).  
CC -!- FUNCTION: Transcription factor that specifies expression of key  
CC genes in developing central nervous system (CNS). Essential for  
CC many, if not all, late developing neuroblastoma (NB) sublineages.  
CC Binds to the 5'-[CG]C[CT][CT]AAAAA[AT]-3' DNA sequence, like hb,  
CC suggesting that cast and hb act as a late regulators in early and  
CC late CNS NB sublineage, respectively. Acts by repressing  
CC expression of nub/pdm-1 and pdm2/pdm-2 POU genes, and restrict  
CC their pattern of expression in appropriate cells. Required for a  
CC full expression of vvl/drifter and acj6/1-POU; it is however  
CC unknown whether it directly activates these genes. Controls  
CC engrailed (en) expression in the ventral nerve chord.  
CC -!- SUBCELLULAR LOCATION: Nucleus.  
CC -!- TISSUE SPECIFICITY: Expressed in a specific subset of neuroblasts  
CC in the ventral nerve chord and the procephalic region in the  
CC embryo. Expressed in many, if not all, late delaminating NBs, and  
CC in early NBs, but only after they have undergone several rounds of  
CC ganglion mother cell-producing divisions.  
CC -!- DEVELOPMENTAL STAGE: Expressed in embryos. Expressed from  
CC blastoderm embryos. Not expressed in first and second instar  
CC larvae. Weakly expressed in third instar larvae. May be weakly  
CC expressed in adults.  
CC -!- SIMILARITY: Contains 1 A.T hook DNA-binding domain.  
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.  
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
CC frameshift in position 782.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: L04487; -; NOT ANNOTATED CDS; mRNA.  
CC EMBL: AE003602; AAF51966.1; -; Genomic\_DNA.  
CC EMBL: AY051865; AAK93289.1; -; mRNA.  
CC EMBL: Z15091; -; NOT ANNOTATED\_CDS; mRNA.  
CC PIR: A49129; A49129.  
CC PIR: JH0797; JH0797.  
CC Ensembl: CG2102; *Drosophila melanogaster*.  
CC FlyBase: FBgn004878; cas.  
CC BioCyc: DMEL-XXX-02:DMEL-XXX-02-010145-MONOMER; -;  
CC BioCyc: DMEL-XXX-02:DMEL-XXX-02-010146-MONOMER; -;  
CC GO: GO:0005634; C:nucleus; IDA.  
CC GO: GO:0003700; P:transcription factor activity; NAS.  
CC GO: GO:0007417; P:central nervous system development; IMP.  
CC GO: GO:0045892; P:negative regulation of transcription, DNA-d. .; IMP.  
CC InterPro: IPR000637; AT hook DNA\_bd.  
CC InterPro: IPR007087; ZnF\_C2H2.  
CC Pfam: PF02178; AT hook; 1.  
CC Pfam: PF00096; zf-C2H2; 1.  
CC SMART: SM00355; ZnF\_C2H2; 3.  
CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
CC PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 2.  
CC Complete proteome; Developmental protein; Differentiation;  
CC DNA-binding; Metal-binding; Neurogenesis; Nuclear protein; Repressor;  
CC Transcription; Transcription regulation; Zinc; Zinc-finger.  
CHAIN 1 793 /FTID=PRO\_0000046911.  
FT ZN\_FING 377 402  
FT C2H2-type 1; atypical.  
FT ZN\_FING 439 463  
FT C2H2-type 2.  
FT ZN\_FING 498 522  
FT C2H2-type 3.  
FT ZN\_FING 556 580  
FT C2H2-type 4.  
FT DNA\_BIND 650 662  
FT A.T hook.  
FT COMPIAS 19 148  
FT Gln-rich.

```

FT COMPIAS 583 625 Gln-rich.
FT COMPIAS 710 713 Poly-Ala.
FT COMPIAS 758 776 Ala-rich.
FT CONFLICT 54 59 R -> E (in Ref. 1).
FT CONFLICT 79 79 R -> A (in Ref. 1).
FT CONFLICT 112 121 A -> G (in Ref. 1).
FT CONFLICT 221 222 G -> R (in Ref. 1).
FT CONFLICT 634 634 S -> T (in Ref. 1).
FT CONFLICT 710 710 A -> R (in Ref. 1).
FT CONFLICT 724 724 Q -> R (in Ref. 1).
SQ SEQUENCE 793 AA; 87925 MW; E35DE18DD3C37671 CRC64;

Query Match 72.2%; Score 39; DB 1; Length 793;
Best Local Similarity 69.2%; Pred. No. 7.2e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ASMSAASAASMAA 13
|:|||||:
Db 758 AALSASAASAAA 770

RESULT 33
Q2UQF0 ASPOR PRELIMINARY; PRT; 1014 AA.
AC Q2UQF0;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, entry version 1.
DE MRNA cleavage and polyadenylation factor II complex.
GN ORFNames=A0090005001277;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=RIB 40;
RX PubMed=16372010; DOI=10.1038/nature04300;
RA Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Nierman W.C., Yu J., Archer N.D.B., Bennett J.W.,
RA Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA Kihara S., Ogasawara N., Kikuchi H.;
RT "Genome sequencing and analysis of Aspergillus oryzae.";
RL Nature 438:1157-1161(2005).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; AP007151; BA556215.1; -; Genomic DNA.
SQ SEQUENCE 1014 AA; 109950 MW; 29FB09A9CAB2F5EA CRC64;

Query Match 72.2%; Score 39; DB 2; Length 1014;
Best Local Similarity 76.9%; Pred. No. 9.3e+02;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAASMAA 13
|:|||||:
Db 124 ASTSAASAASAAA 136

RESULT 34
Q63LK9_BURPS PRELIMINARY; PRT; 4574 AA.
ID Q63LK9_BURPS
AC Q63LK9;
DT 25-OCT-2004, integrated into UniProtKB/TrEMBL.

```

```

DT 25-OCT-2004, sequence version 1.
DE Putative polyketide synthase.
GN OrderedLocusNames=BPSS1006;
OS Burkholderia pseudomallei (Pseudomonas pseudomallei).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia; pseudomallei group.
OX NCBI_TaxID=28450;
RN [1];
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=K96243;
RX PubMed=15377794; DOI=10.1073/pnas.0403302101;
RA Holden M.T.G., Titball R.W., Peacock S.J., Cerdeno-Tarraga A.-M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.L.,
RA Bentley S.D., Sebaihia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crosssett B., Davis P., Deshazer D.,
RA Feltwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitsch E., Rutherford K., Sanders M., Simmonds M.,
RA Songvilai S., Stevens K., Tumapa S., Vesaratchaveat M.,
RA Whitehead S., Yeats C., Barrell B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei".
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
DR EMBL; BX571966; CAH38467.1; -; Genomic DNA.
DR GO; GO:0048037; F:cofactor binding; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006520; P:amino acid metabolism; IEA.
DR GO; GO:0006533; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR009081; ACP like.
DR InterPro; IPR001597; Beta_elim_lyase.
DR InterPro; IPR000794; Ketoacyl_synth.
DR InterPro; IPR006163; Phosphopanteth_bd.
DR InterPro; IPR006162; Ppantne_S.
DR InterPro; IPR002198; SDR.
DR PANTHER; PTHR11712; Ketoacyl_synth; 8.
DR Pfam; PF00106; adh short; 1.
DR Pfam; PF01212; Beta_elim_lyase; 2.
DR Pfam; PF00109; ketoacyl_synth; 3.
DR Pfam; PF02801; ketoacyl_synth_C; 3.
DR Pfam; PF00550; PP-binding; 2.
DR ProDom; PD005927; Beta_elim_lyase; 1.
DR PROSITE; PS00075; ACP DOMAIN; 2.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_2.
KW Complete proteome.
SQ SEQUENCE 4574 AA; 484884 MW; F8FA0909D2D9C75F CRC64;

Query Match 72.2%; Score 39; DB 2; Length 4574;
Best Local Similarity 76.9%; Pred. No. 4.2e+03;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ASMSAASAASMAA 13
|:|||||:
Db 3959 ASASAASAASAAAS 3971

RESULT 35
Q2QLR1_ORYSA PRELIMINARY; PRT; 129 AA.
ID Q2QLR1_ORYSA
AC Q2QLR1;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Hypothetical protein.
GN ORFNames=LOC_0612943610;

```

```
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP Clade;
OC Eriartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Wing R.A., McCombie W.A., Ouyang S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DP000011; ABA99936.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 129 AA; 13507 MW; 61F89A3E8FF8AA49 CRC64;

Query Match 70.4%; Score 38; DB 2; Length 129;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASASASMA 12
Db |||:||||:|
46 ASSSASASASLA 57

RESULT 36
Q84JT2 ORYSA
ID Q84JT2 ORYSA PRELIMINARY; PRT; 166 AA.
AC Q84JT2;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, sequence version 1.
DT 21-FEB-2006, entry version 13.
DE Hypothetical protein OSJNB0096M04.140 (Hypothetical protein
DE OSJNB0043P23.4).
GN Name=OSJNB0096M04.140; Synonyms=OSJNB0043P23.4;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP Clade;
OC Eriartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Uterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RA "Oryza sativa chromosome 3 BAC OSJNB0096M04 genomic sequence.";
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Uterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RA "Oryza sativa chromosome 3 BAC OSJNB0043P23 genomic sequence.";
RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
```

```
DR EMBL; AC092559; AAO37944.1; -; Genomic_DNA.
DR EMBL; AC099324; AAO73278.1; -; Genomic_DNA.
DR Gramene; Q84JT2;
DR InterPro; IPR000095; PAK_box_Rho_bd.
DR Pfam; PF00786; PBD; 1.
DR PROSITE; PS50108; CRIB; 1.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 17181 MW; 6EC080E8429A16DB CRC64;

Query Match 70.4%; Score 38; DB 2; Length 166;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 SMSAASASASMA 13
Db |||:||||:|
42 SSSASASASMA 53

RESULT 37
Q2UKI7 ASPOR
ID Q2UKI7 ASPOR PRELIMINARY; PRT; 169 AA.
AC Q2UKI7;
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 3.
DE Predicted protein.
GN ORFNames=AO090003000790;
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eriartoideae; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buchwald M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,
RA Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,
RA Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,
RA Galagan J.E., Niezman W.C., Yu J., Archer D.B., Bennett J.W.,
RA Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotoh O., Horikawa H.,
RA Hosoyama A., Ichinomiya M., Igarashi R., Iwashita K., Juvvadi P.R.,
RA Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,
RA Maruyama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,
RA Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,
RA Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,
RA Komori T., Koyama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,
RA Kuhara S., Ogasawara N., Kikuchi H.;
RA "Genome sequencing and analysis of Aspergillus oryzae.";
RL Nature 438:1157-1161(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP007155; BAE57928.1; -; Genomic_DNA.
SQ SEQUENCE 169 AA; 17228 MW; 3A2A8F472B04F675 CRC64;

Query Match 70.4%; Score 38; DB 2; Length 169;
Best Local Similarity 69.2%; Pred. No. 2.2e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASASMA 13
Db |||:||||:|
8 ASLAASASASMA 20

RESULT 38
Q2XY88 DROYA
ID Q2XY88 DROYA PRELIMINARY; PRT; 172 AA.
AC Q2XY88;
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 20-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE CG13448 (Fragment).
```

```
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T33;
RX PubMed=16120803; DOI=10.1093/molbev/msi246;
RA Comeron J.M., Guthrie T.B.;
RT "Intragenic Hill-Robertson Interference Influences Selection Intensity
on Synonymous Mutations in Drosophila.";
RL Mol. Biol. Evol. 22:2519-2530(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DQ138774; ABA86380.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 172 172
SQ SEQUENCE 172 AA; 17735 MW; CF32B931EFA888DE CRC64;

Query Match 70.4%; Score 38; DB 2; Length 172;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SMSAASAASMAA 13
Db 51 SYSAATAAAMAA 62

RESULT 39
Q2XY89 DROYA
ID Q2XY89 DROYA PRELIMINARY; PRT; 172 AA.
AC Q2XY89;
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE CGI3448 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Tail8;
RX PubMed=16120803; DOI=10.1093/molbev/msi246;
RA Comeron J.M., Guthrie T.B.;
RT "Intragenic Hill-Robertson Interference Influences Selection Intensity
on Synonymous Mutations in Drosophila.";
RL Mol. Biol. Evol. 22:2519-2530(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DQ138773; ABA86379.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 172 172
SQ SEQUENCE 172 AA; 17739 MW; D2837931EFA88D9B CRC64;

Query Match 70.4%; Score 38; DB 2; Length 172;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SMSAASAASMAA 13
Db 51 SYSAATAAAMAA 62

RESULT 40
Q2XY87 DROER
ID Q2XY87_DROER PRELIMINARY; PRT; 174 AA.
```

```
AC Q2XY87;
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 1.
DE CGI3448 (Fragment).
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=16120803; DOI=10.1093/molbev/msi246;
RA Comeron J.M., Guthrie T.B.;
RT "Intragenic Hill-Robertson Interference Influences Selection Intensity
on Synonymous Mutations in Drosophila.";
RL Mol. Biol. Evol. 22:2519-2530(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; DQ138775; ABA86381.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 174 174
SQ SEQUENCE 174 AA; 17861 MW; 76B5A0CDB39E7C7E CRC64;

Query Match 70.4%; Score 38; DB 2; Length 174;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SMSAASAASMAA 13
Db 51 SYSAATAAAMAA 62

Search completed: September 9, 2006, 23:00:14
Job time : 125.139 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:37:17 ; Search time 97.4177 Seconds  
(without alignments)  
61.014 Million cell updates/sec

Title: US-10-617-568-36  
Perfect score: 44  
Sequence: 1 AAXAAAAAXAA 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A Geneseq\_8:\*

1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*  
9: Geneseqp2005s: \*  
10: Geneseqp2006s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	13	9	Adv78640 Cell atta
2	44	100.0	13	10	Aed86538 Silk pept
3	44	100.0	14	9	Adv78641 Cell atta
4	44	100.0	14	10	Aed86539 Silk pept
5	44	100.0	15	6	ABR38361 Human can
6	44	100.0	15	6	ABR38363 Human can
7	44	100.0	15	6	ABR38360 Human can
8	44	100.0	15	6	ABR38364 Human can
9	44	100.0	15	6	ABR38365 Human can
10	44	100.0	15	6	ABR38362 Human can
11	44	100.0	15	8	ADO58104 Spider be
12	44	100.0	15	9	Adv78642 Cell atta
13	44	100.0	15	10	Aed86540 Silk pept
14	44	100.0	15	10	Aef42853 Arg-C pro
15	44	100.0	16	2	AAW32224 Beta-shee
16	44	100.0	16	2	AAW32226 Beta-shee
17	44	100.0	16	9	Adv78643 Cell atta
18	44	100.0	17	2	AAW32227 Beta-shee
19	44	100.0	18	7	AAO30390 Human ARX
20	44	100.0	18	8	ADH58940 Silk prot
21	44	100.0	18	10	Aef40057 Polyvalani
22	44	100.0	19	7	AAO30392 Human ARX
23	44	100.0	19	9	Adv78587 Cell atta

24	44	100.0	19	9	AEB23102
25	44	100.0	20	4	AAB69489
26	44	100.0	20	7	ADH47641 Human lun
27	44	100.0	20	8	ADH58955 Silk prot
28	44	100.0	20	8	ADJ21560 Human lun
29	44	100.0	20	9	Adv78588 Cell atta
30	44	100.0	20	10	Aef40072 Domestic
31	44	100.0	21	2	AAI25436
32	44	100.0	21	4	AAB69490
33	44	100.0	21	7	ADJ81212 Self-as
34	44	100.0	21	7	ADJ81342 Self-as
35	44	100.0	21	9	Adv78589 Cell atta
36	44	100.0	22	9	Adv78590 Cell atta
37	44	100.0	23	9	Adv78591 Cell atta
38	44	100.0	24	9	Adv78592 Cell atta
39	44	100.0	25	5	ABG95821 Transmem
40	44	100.0	26	7	ADJ81343 Self-as
41	44	100.0	26	7	ADJ81213 Self-as
42	44	100.0	26	9	AEB23106 Cell adhe
43	44	100.0	27	10	Aef42855 Asp-N end
44	44	100.0	28	7	ADJ81344 Self-as
45	44	100.0	28	7	ADJ81214 Self-as
46	44	100.0	28	9	AEB23104 Cell adhe
47	44	100.0	28	9	AEB23105 Cell adhe
48	44	100.0	31	3	AAB08166 Peptide m
49	44	100.0	32	9	ADH83853 Heparin b
50	44	100.0	38	9	Adv78600 Cell atta
51	44	100.0	39	5	ABG95847 Fusion pe
52	44	100.0	40	9	Adv78601 Cell atta
53	44	100.0	40	9	AEB23103 Cell adhe
54	44	100.0	40	9	AED01242 Wound dre
55	44	100.0	40	9	AEE05648 Auxiliary
56	44	100.0	40	9	AEE01614 Auxilliari
57	44	100.0	40	10	Aef42850 Asp-N/Arg
58	44	100.0	40	10	Aef82116 Cell cult
59	44	100.0	42	9	Adv78602 Cell atta
60	44	100.0	44	9	Adv78603 Cell atta
61	44	100.0	46	9	Adv78604 Cell atta
62	44	100.0	48	9	Adv78605 Cell atta
63	44	100.0	50	5	ABG95846 Fusion pe
64	44	100.0	50	5	ABG95834 Fusion pe
65	44	100.0	66	6	AAB36900 Spider (P
66	44	100.0	67	7	ADI40644 Anti-coag
67	44	100.0	68	7	ADI40633 Anti-coag
68	44	100.0	78	7	ADE10614 Structura
69	44	100.0	78	7	ADE10664 Structura
70	44	100.0	78	8	ADK15683 Library f
71	44	100.0	78	8	ADK15633 Nucleatin
72	44	100.0	79	7	ADE10615 Structura
73	44	100.0	79	7	ADE10665 Structura
74	44	100.0	79	8	ADK15634 Nucleatin
75	44	100.0	79	8	ADK15684 Library f
76	44	100.0	79	9	ADY27669 Human ach
77	44	100.0	88	7	ADE10617 Structura
78	44	100.0	88	7	ADE10667 Structura
79	44	100.0	88	8	ADK15636 Nucleatin
80	44	100.0	88	8	ADK15686 Library f
81	44	100.0	92	8	ADN13984 Human pro
82	44	100.0	92	10	Aef29458 Lead Cere
83	44	100.0	96	4	ABG16827 Novel hum
84	44	100.0	96	7	ADE10668 Structura
85	44	100.0	96	7	ADE10618 Structura
86	44	100.0	96	8	ADK15637 Nucleatin
87	44	100.0	96	8	ADK15687 Library f
88	44	100.0	98	2	AAI25428 Antifreeze
89	44	100.0	98	2	AAI25428 Synthetic
90	44	100.0	102	6	ABJ20256 162p1B6 c
91	44	100.0	102	6	ABJ20253 162p1B6 c
92	44	100.0	102	6	ABJ20243 162p1B6 c
93	44	100.0	104	7	ADCE4293 Human SID
94	44	100.0	104	7	ADCE4293 Human SID
95	44	100.0	104	7	ADE10628 Structura
96	44	100.0	104	7	ADE10629 Structura

97 44 100.0 104 7 ADE10624 Structura  
98 44 100.0 104 7 ADE10677 Structura  
99 44 100.0 104 7 ADE10681 Structura  
100 44 100.0 104 7 ADE10626 Structura

## ALIGNMENTS

RESULT 1  
ADV78640  
ID ADV78640 standard; peptide; 13 AA.  
XX  
AC ADV78640;  
XX  
XX 24-MAR-2005 (first entry)  
XX  
DE Cell attachment-related auxillary peptide - SEQ ID 104.  
XX  
XX cell attachment; cell culture; transplantation; vaccine; antibiotic.  
XX  
OS Synthetic.  
XX  
XX JP2005002106-A.  
PN  
XX 06-JAN-2005.  
PD  
XX  
XX 20-MAY-2004; 2004JP-00150510.  
PF  
XX  
XX 21-MAY-2003; 2003JP-00144125.  
PR  
XX  
XX (SANN ) SANYO CHEM IND LTD.  
PA  
XX  
XX WPI; 2005-053150/06.  
DR  
XX  
XX Cell attachable polypeptide useful for culturing cell, comprises  
PT auxiliary amino acid sequence having specific amount of Gly and Ala,  
PT chemically bonded to cell attachable minimal amino acid sequence,  
PT alternatively.  
XX  
XX Disclosure; SEQ ID NO 104; 20pp; Japanese.

XX The invention relates to a novel cell attachable polypeptide comprising  
XX an auxiliary amino acid sequence (Y) chemically bonded to a cell  
CC attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino  
CC acids and the sum total content of Gly and Ala in (Y) is 42-100% with  
CC respect to the total number of amino acids in (Y). The polypeptide of the  
CC invention may be utilized during culturing of cells for use in various  
CC experiments, including toxicity studies, gene transfer and production of  
CC useful substances such as cytokines, thrombolytic agents, vaccines,  
CC hormones and antibiotics. Furthermore, the polypeptide may be useful for  
CC culturing cells for transplantation, particularly, transplantation of  
CC skin, cranial bones, muscles, skin tissue, bone marrow and retina. The  
CC polypeptide exhibits extremely potent cell adhesion properties and can  
CC attach cells, such as NHDF and VERO cells, more efficiently onto a base  
CC material, relative to conventional methods. The polypeptide also exhibits  
CC high heat stability, thus sterilization in autoclaves can be performed  
CC conveniently. The current sequence is that of a cell attachment-related  
CC auxiliary peptide of the invention.

XX Sequence 13 AA;

Query Match 100.0%; Score 44; DB 9; Length 13;  
Best Local Similarity 84.6%; Pred. No. 1.8;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 1 AAAAAAAAAXAA 13

RESULT 2  
AED86538

ID AED86538 standard; peptide; 13 AA.  
XX  
AC AED86538;  
XX  
DT 26-JAN-2006 (first entry)  
XX  
DE Silk peptide linker, SEQ ID 10.  
XX  
KW Silk.  
XX  
OS Nephila clavipes.  
XX  
PN US2005261425-A1.  
XX  
PD 24-NOV-2005.  
XX  
XX 25-APR-2005; 2005US-00113494.  
PF  
XX 13-MAY-2002; 2002US-0380502P.  
PR  
XX 13-MAY-2003; 2003US-00436611.  
XX  
XX (CORR ) CORNELL RES FOUND INC.  
PA  
XX Rathore O, Sogah DY;  
PI  
XX WPI; 2006-008467/01.  
DR  
XX  
XX Composition useful for producing fibers or films, comprises hybrid  
PT multiblock copolymer derived from Nephila clavipes or Bombyx mori silks.  
XX  
XX Disclosure; SEQ ID NO 10; 27pp; English.  
XX  
XX The present invention relates to a novel composition (C1) comprising a  
CC hybrid multiblock copolymer. The multiblock copolymer comprise peptide  
CC linkers, which are derived from peptide sequences found in native silk  
CC biomacromolecules or the crystalline domains of Nephila clavipes or  
CC Bombyx mori silks. The present sequence is one such peptide linker. (C1)  
CC is useful for producing films or fibers utilized in tissue-engineering  
CC scaffolds, cell culture substrates, biocompatible coatings, sutures,  
CC membranes, grafts and drug delivery systems. (C1) enables to produce  
CC fibers or films having increased stiffness and tensile strength.  
XX  
XX Sequence 13 AA;  
SQ  
Query Match 100.0%; Score 44; DB 10; Length 13;  
Best Local Similarity 84.6%; Pred. No. 1.8;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAXAAAAAAAXAA 13  
||| ||||| ||  
Db 1 AAAAAAAAAXAA 13  
RESULT 3  
ADV78641  
ID ADV78641 standard; peptide; 14 AA.  
XX  
AC ADV78641;  
XX  
XX 24-MAR-2005 (first entry)  
DT  
XX  
XX Cell attachment-related auxillary peptide - SEQ ID 105.  
DE  
XX cell attachment; cell culture; transplantation; vaccine; antibiotic.  
KW  
XX Synthetic.  
OS  
XX JP2005002106-A.  
PN  
XX 06-JAN-2005.  
PD  
XX  
XX 20-MAY-2004; 2004JP-00150510.  
PF  
XX





CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
CC from the invention

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 6; Length 15;  
Best Local Similarity 84.6%; Pred. No. 2;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
|| ||||| ||  
Db 2 AAAAAAAAAXAA 14

RESULT 6  
ABR38363  
ID ABR38363 standard; peptide; 15 AA.

AC ABR38363;

XX 19-MAY-2003 (first entry)

XX Human cancer-related protein 187P3F2 HLA peptide #1494.

XX Human; cytostatic; vaccine; cancer; immune response; HLA;  
KW human leukocyte antigen.

XX Homo sapiens.

XX WO200283921-A2.

XX 24-OCT-2002.

XX 10-APR-2002; 2002WO-US011654.

XX 10-APR-2001; 2001US-0282739P.

XX 10-APR-2001; 2001US-0283112P.

XX 25-APR-2001; 2001US-0286630P.

XX (AGEN-) AGENSYS INC.

XX Jakobovits A, Challita-Eid PM, Paris M, Ge W, Hubert RS;  
PI Morrison K, Morrison RK, Raitano AB;

XX WPI; 2003-075555/07.

XX New composition comprising a substance that modulates the structure of  
PT proteins and polynucleotides, useful for therapeutic, prognostic and  
PT diagnostic reagents for eliciting cellular or humoral immune response in  
PT cancer patients.

XX Claim 13; Page 618; 1021pp; English.

XX The present invention relates to novel human cancer-related genes and  
CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
CC proteins are useful for eliciting a humoral or cellular immune response.  
CC The genes are useful as probes and primers for the amplification and/or  
CC detection of genes, mRNAs or their fragments, as reagents for the  
CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
CC directing the expression of the protein, as tools for modulating or  
CC inhibiting the expression of genes and/or translation of transcripts, and  
CC as therapeutic agents. The proteins and peptides are useful as  
CC therapeutic, prognostic and diagnostic reagents for cancer. The present  
CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
CC from the invention

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 6; Length 15;  
Best Local Similarity 84.6%; Pred. No. 2;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13

Db || ||||| ||  
1 AAAAAAAAAXAA 13

RESULT 7  
ABR38360

ID ABR38360 standard; peptide; 15 AA.

XX ABR38360;

XX 19-MAY-2003 (first entry)

XX Human cancer-related protein 187P3F2 HLA peptide #1491.

XX Human; cytostatic; vaccine; cancer; immune response; HLA;  
KW human leukocyte antigen.

XX Homo sapiens.

XX WO200283921-A2.

XX 24-OCT-2002.

XX 10-APR-2002; 2002WO-US011654.

XX 10-APR-2001; 2001US-0282739P.

XX 10-APR-2001; 2001US-0283112P.

XX 25-APR-2001; 2001US-0286630P.

XX (AGEN-) AGENSYS INC.

XX Jakobovits A, Challita-Eid PM, Paris M, Ge W, Hubert RS;  
PI Morrison K, Morrison RK, Raitano AB;

XX WPI; 2003-075555/07.

XX New composition comprising a substance that modulates the structure of  
PT proteins and polynucleotides, useful for therapeutic, prognostic and  
PT diagnostic reagents for eliciting cellular or humoral immune response in  
PT cancer patients.

XX Claim 13; Page 618; 1021pp; English.

XX The present invention relates to novel human cancer-related genes and  
CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and  
CC proteins are useful for eliciting a humoral or cellular immune response.  
CC The genes are useful as probes and primers for the amplification and/or  
CC detection of genes, mRNAs or their fragments, as reagents for the  
CC diagnosis and/or prognosis of cancer, as coding sequences capable of  
CC directing the expression of the protein, as tools for modulating or  
CC inhibiting the expression of genes and/or translation of transcripts, and  
CC as therapeutic agents. The proteins and peptides are useful as  
CC therapeutic, prognostic and diagnostic reagents for cancer. The present  
CC sequence is a human leukocyte antigen (HLA) peptide, used in an example  
CC from the invention

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 6; Length 15;  
Best Local Similarity 84.6%; Pred. No. 2;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13

Db 3 AAAAAAAAAXAA 15

RESULT 8

ABR38364

ID ABR38364 standard; peptide; 15 AA.

XX ABR38364;

XX

```

DT XX 19-MAY-2003 (first entry)
DE XX Human cancer-related protein 187P3F2 HLA peptide #1495.
KW KW Human; cytostatic; vaccine; cancer; immune response; HLA;
XX KW human leukocyte antigen.
OS XX Homo sapiens.
XX OS
XX PN WO200283921-A2.
XX XX
XX PD 24-OCT-2002.
XX XX
XX PF 10-APR-2002; 2002WO-US011654.
XX XX
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 10-APR-2001; 2001US-0283112P.
XX PR 25-APR-2001; 2001US-0286630P.
XX XX
XX PA (AGEN-) AGENSYS INC.
XX XX
XX PI Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;
PI PI Morrison K, Morrison RK, Raitano AB;
XX XX
XX DR WPI; 2003-075555/07.
XX XX
XX PT New composition comprising a substance that modulates the structure of
XX PT proteins and polynucleotides, useful for therapeutic, prognostic and
XX PT diagnostic reagents for eliciting cellular or humoral immune response in
XX PT cancer patients.
XX XX
XX PS Claim 13; Page 618; 1021pp; English.
XX XX
XX CC The present invention relates to novel human cancer-related genes and
XX CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and
XX CC proteins are useful for eliciting a humoral or cellular immune response.
XX CC The genes are useful as probes and primers for the amplification and/or
XX CC detection of genes, mRNAs or their fragments, as reagents for the
XX CC diagnosis and/or prognosis of cancer, as coding sequences capable of
XX CC directing the expression of the protein, as tools for modulating or
XX CC inhibiting the expression of genes and/or translation of transcripts, and
XX CC as therapeutic agents. The proteins and peptides are useful as
XX CC therapeutic, prognostic and diagnostic reagents for cancer. The present
XX CC sequence is a human leukocyte antigen (HLA) peptide, used in an example
XX CC from the invention
XX CC
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 44; DB 6; Length 15;
XX Best Local Similarity 84.6%; Pred. No. 2;
XX Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 AAXAAAAAAAXAA 13
Db || || || || || || ||
1 AAAAAAAAAXAA 13

RESULT 9
ABR38365
ID ABR38365 standard; peptide; 15 AA.
XX XX
XX AC ABR38365;
XX XX
XX DT 19-MAY-2003 (first entry)
XX XX
XX DE Human cancer-related protein 187P3F2 HLA peptide #1496.
XX KW Human; cytostatic; vaccine; cancer; immune response; HLA;
XX KW human leukocyte antigen.
XX OS Homo sapiens.
XX OS
XX PN WO200283921-A2.
XX XX

DT XX 19-MAY-2003 (first entry)
DE XX Human cancer-related protein 187P3F2 HLA peptide #1493.
KW KW Human; cytostatic; vaccine; cancer; immune response; HLA;
XX KW human leukocyte antigen.
XX OS Homo sapiens.
XX OS
XX PN WO200283921-A2.
XX XX
XX PD 24-OCT-2002.
XX XX
XX PF 10-APR-2002; 2002WO-US011654.
XX XX
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 10-APR-2001; 2001US-0283112P.
XX PR 25-APR-2001; 2001US-0286630P.
XX XX
XX PA (AGEN-) AGENSYS INC.
XX XX
XX PI Jakobovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;
PI PI Morrison K, Morrison RK, Raitano AB;
XX XX
XX DR WPI; 2003-075555/07.
XX XX
XX PT New composition comprising a substance that modulates the structure of
XX PT proteins and polynucleotides, useful for therapeutic, prognostic and
XX PT diagnostic reagents for eliciting cellular or humoral immune response in
XX PT cancer patients.
XX XX
XX PS Claim 13; Page 618; 1021pp; English.
XX XX
XX CC The present invention relates to novel human cancer-related genes and
XX CC proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and
XX CC proteins are useful for eliciting a humoral or cellular immune response.
XX CC The genes are useful as probes and primers for the amplification and/or
XX CC detection of genes, mRNAs or their fragments, as reagents for the
XX CC diagnosis and/or prognosis of cancer, as coding sequences capable of
XX CC directing the expression of the protein, as tools for modulating or
XX CC inhibiting the expression of genes and/or translation of transcripts, and
XX CC as therapeutic agents. The proteins and peptides are useful as
XX CC therapeutic, prognostic and diagnostic reagents for cancer. The present
XX CC sequence is a human leukocyte antigen (HLA) peptide, used in an example
XX CC from the invention
XX CC
XX SQ Sequence 15 AA;
XX
XX Query Match 100.0%; Score 44; DB 6; Length 15;
XX Best Local Similarity 84.6%; Pred. No. 2;
XX Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 AAXAAAAAAAXAA 13
Db || || || || || || ||
1 AAAAAAAAAXAA 13

RESULT 10
ABR38362
ID ABR38362 standard; peptide; 15 AA.
XX XX
XX AC ABR38362;
XX XX
XX DT 19-MAY-2003 (first entry)
XX XX
XX DE Human cancer-related protein 187P3F2 HLA peptide #1493.
XX XX
XX KW Human; cytostatic; vaccine; cancer; immune response; HLA;
XX KW human leukocyte antigen.
XX OS Homo sapiens.
XX OS
XX PN WO200283921-A2.
XX XX
XX PD 24-OCT-2002.
XX XX
XX PF 10-APR-2002; 2002WO-US011654.
XX XX
XX PR 10-APR-2001; 2001US-0282739P.
XX PR 10-APR-2001; 2001US-0283112P.
XX PR 25-APR-2001; 2001US-0286630P.
XX XX
XX PA (AGEN-) AGENSYS INC.
XX XX

```

XX	Jakovovits A, Challita-Eid PM, Faris M, Ge W, Hubert RS;
PI	Morrison K, Morrison RK, Raitano AB;
XX	WPI; 2003-075555/07.
XX	New composition comprising a substance that modulates the structure of
PT	proteins and polynucleotides, useful for therapeutic, prognostic and
PT	diagnostic reagents for eliciting cellular or humoral immune response in
PT	cancer patients.
XX	Claim 13; Page 618; 1021pp; English.
PS	The present invention relates to novel human cancer-related genes and
XX	proteins (ABZ78120-ABZ78168 and ABR01789-ABR01861). The genes and
CC	proteins are useful for eliciting a humoral or cellular immune response.
CC	The genes are useful as probes and primers for the amplification and/or
CC	detection of genes, mRNAs or their fragments, as reagents for the
CC	diagnosis and/or prognosis of cancer, as coding sequences capable of
CC	directing the expression of the protein, as tools for modulating or
CC	inhibiting the expression of genes and/or translation of transcripts, and
CC	as therapeutic agents. The proteins and peptides are useful as
CC	therapeutic, prognostic and diagnostic reagents for cancer. The present
CC	sequence is a human leukocyte antigen (HLA) peptide, used in an example
CC	from the invention
XX	Sequence 15 AA;
XX	Query Match 100.0%; Score 44; DB 6; Length 15;
XX	Best Local Similarity 84.6%; Pred. No. 2;
XX	Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0
QY	1 AAXXXXXXXXXAA 13
DB	1 AAAAAAAAAAAAA 13
RESULT 11	
ADO58104	
ID	ADO58104 standard; peptide; 15 AA.
XX	ADO58104;
XX	12-AUG-2004 (first entry)
DT	Spider beta-sheet crystalline segment found in silk biopolymer.
DE	Synthetic multiblock copolymer; linker; peptide block;
XX	native silk biomacromolecule; crystalline domain; structural biopolymer;
KW	silk; beta-sheet forming crystalline segment; film; fibre;
KW	high tensile strength; spider; beta-sheet crystalline segment.
XX	Nephila clavipes.
OS	US2004102608-A1.
XX	27-MAY-2004.
PN	13-MAY-2003; 2003US-00436611.
XX	13-MAY-2002; 2002US-0380502P.
XX	(CORR ) CORNELL RES FOUND INC.
PA	Rathore O, Sogah DY;
XX	WPI; 2004-447966/42.
XX	Multiblock copolymer useful for producing fibers and films having high
PT	tensile strength, comprising linkers and peptide blocks derived from
PT	peptide sequences found in native silk biomacromolecules of Nephila
PT	clavipes.
XX	

PS Disclosure; SEQ ID NO 3; 23pp; English.

XX

CC The present invention relates to synthetic multiblock copolymers comprising linkers and peptide blocks derived from peptide sequences found in native silk biomacromolecules such as the crystalline domain of the spider, Nephila clavipes, or the silkworm, Bombyx mori. The synthetic multiblock copolymers are produced by replacing the amorphous peptide domain of a structural biopolymer such as silk with a nonpeptide segment while maintaining the beta-sheet forming crystalline segments. The multiblock copolymers have solid-state structures and mechanical properties similar to the naturally occurring structural biopolymers. The copolymers are useful for producing films or fibres with high tensile strength. The present sequence represents a spider beta-sheet crystalline segment found in the silk biopolymer.

XX

SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 8; Length 15;  
Best Local Similarity 84.6%; Pred.No. 2;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAAAAXAA 13  
|| || || || || || || ||  
DB 1 AAAAAAAXAAAA 13

RESULT 12

ID ADV78642

XX ADV78642 standard; peptide; 15 AA.

AC ADV78642;

DT 24-MAR-2005 (first entry)

XX Cell attachment-related auxillary peptide - SEQ ID 106.  
cell attachment; cell culture; transplantation; vaccine; antibiotic.  
Synthetic.

OS JP2005002106-A.

PN 06-JAN-2005.

PD 20-MAY-2004; 2004JP-00150510.

PF 21-MAY-2003; 2003JP-00144125.

PR (SANN ) SANYO CHEM IND LTD...  
WPI; 2005-053150/06.

XX Cell attachable polypeptide useful for culturing cell, comprises auxiliary amino acid sequence having specific amount of Gly and Ala, chemically bonded to cell attachable minimal amino acid sequence, alternatively.

PT

XX Disclosure; SEQ ID NO 106; 20pp; Japanese.

XX

CC The invention relates to a novel cell attachable polypeptide comprising an auxiliary amino acid sequence (Y) chemically bonded to a cell attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino acids and the sum total content of Gly and Ala in (Y) is 42-100% with respect to the total number of amino acids in (Y). The polypeptide of the invention may be utilized during culturing of cells for use in various experiments, including toxicity studies, gene transfer and production of useful substances such as cytokines, thrombolytic agents, vaccines, hormones and antibiotics. Furthermore, the polypeptide may be useful for culturing cells for transplantation, particularly, transplantation of skin, cranial bones, muscles, skin tissue, bone marrow and retina. The polypeptide exhibits extremely potent cell adhesion properties and can attach cells, such as NHDF and VERO cells, more efficiently onto a base material, relative to conventional methods. The polypeptide also exhibits

CC high heat stability, thus sterilization in autoclaves can be performed  
 CC conveniently. The current sequence is that of a cell attachment-related  
 CC auxillary peptide of the invention.

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 9; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 2;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAXAA 13  
 |||||  
 Db 1 AAAAAAXAA 13

RESULT 13  
 AED86540  
 ID AED86540 standard; peptide; 15 AA.

AC AED86540;

XX 26-JAN-2006 (first entry)

XX Silk peptide linker, SEQ ID 12.

XX Silk.

XX Nephila clavipes.

XX US2005261425-A1.

XX 24-NOV-2005.

XX 25-APR-2005; 2005US-00113494.

XX 13-MAY-2002; 2002US-0380502P.

XX 13-MAY-2003; 2003US-00436611.

XX (CORR ) CORNELL RES FOUND INC.

XX Rathore O, Sogah DY;

XX WPI; 2006-008467/01.

XX Composition useful for producing fibers or films, comprises hybrid  
 PT multiblock copolymer derived from Nephila clavipes or Bombyx mori silks.

XX Disclosure; SEQ ID NO 12; 27pp; English.

XX The present invention relates to a novel composition (C1) comprising a  
 CC hybrid multiblock copolymer. The multiblock copolymer comprise peptide  
 CC linkers, which are derived from peptide sequences found in native silk  
 CC biomacromolecules or the crystalline domains of Nephila clavipes or  
 CC Bombyx mori silks. The present sequence is one such peptide linker. (C1)  
 CC is useful for producing films or fibers utilized in tissue-engineering  
 CC scaffolds, cell culture substrates, biocompatible coatings, sutures,  
 CC membranes, grafts and drug delivery systems. (C1) enables to produce  
 CC fibers or films having increased stiffness and tensile strength.

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 10; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 2;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAXAA 13  
 |||||  
 Db 1 AAAAAAXAA 13

RESULT 14

AEF42853

ID AEF42853 standard; peptide; 15 AA.

XX AEF42853;  
 XX 23-MAR-2006 (first entry)  
 XX Arg-C proteinase digest peptide SEQ ID NO:4.  
 XX Protein degradation; protein detection.  
 XX Synthetic.

XX WO2006005622-A1.

XX 19-JAN-2006.

XX 14-JUL-2005; 2005WO-EP007684.

XX 15-JUL-2004; 2004EP-00016696.

XX (BIOV-) BIOVISION AG.

XX Tammen H, Schulz-Knappe P, Lamping N, Zucht H, Hess R;  
 XX Kellmann M, Menzel C, Schulte I;

XX WPI; 2006-109952/11.

XX Determining, monitoring and/or controlling the quality of a sample  
 CC comprising at least one peptide and/or protein comprises determining the  
 CC level of proteolytic cleavage and/or modification, and/or peptide loss of  
 CC the standard.

XX Example 6; SEQ ID NO 4; 54pp; English.

XX The invention relates to determining, monitoring and/or controlling the  
 CC quality of a sample comprising at least one peptide and/or protein,  
 CC comprising providing at least one protease-sensitive standard and/or at  
 CC least one modification-sensitive standard and/or at least one standard  
 CC for monitoring peptide loss, adding sample comprising at least one  
 CC peptide and/or protein, and determining the level of proteolytic cleavage  
 CC and/or modification and/or peptide loss of the standard. The proteolytic  
 CC cleavage is due to the presence of at least one protease selected from  
 CC serine proteases, cysteine proteases, threonine proteases, aspartic  
 CC proteases, glutamic proteases, thiol proteases, metallo proteases,  
 CC exoproteases, endoproteases and proteases involved in the clotting  
 CC cascade and proteases involved in the classical and/or alternative  
 CC complement cascade. The protease sensitive standard is at least one  
 CC peptide comprising at least one cleavage site for at least one protease.  
 CC The method is useful for determining, monitoring and/or controlling the  
 CC quality of a sample comprising at least one peptide and/or protein,  
 CC preferably for determining the progression of proteolytic cleavage and/or  
 CC modification and/or peptide loss of the standard with time. The present  
 CC sequence is an Arg-C digest peptide from a hypothetical standard peptide  
 CC (containing an Asp-N endoprotease site and an Arg-C proteinase  
 CC (Clostripain and trypsin) site) useful in the method of the invention.

XX SQ Sequence 15 AA;

Query Match 100.0%; Score 44; DB 10; Length 15;  
 Best Local Similarity 84.6%; Pred. No. 2;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAXAA 13  
 |||||  
 Db 1 AAAAAAXAA 13

RESULT 15

AAW32224

ID AAW32224 standard; peptide; 16 AA.

XX AAW32224;

XX 12-FEB-1998 (first entry)

```

XX DE Beta-sheet forming peptide #1.
XX KW Beta sheet; circular dichroism spectroscopy; glycosidase activity;
XX KW phosphodiesterase activity; drug screening; hydrolysis;
XX KW neurodegenerative disease; Alzheimer's disease; amyloid deposition.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /note= "N-terminal acetyl"
XX FT Modified-site 16
XX FT /note= "C-terminal amide"
XX PN W09637212-A1.
XX XX 28-NOV-1996.
XX PD 23-MAY-1996; 96WO-US007564.
XX PF 26-MAY-1995; 95US-00452043.
XX PR (TORR-) TORREY PINES INST MOLECULAR STUDIES.
XX PA Blondelle SE, Perez-Paya E, Forood B, Houghten RA;
XX PI WPI; 1997-033944/03.
XX DR Polypeptide(s) which form stable beta sheets in aq. environment - having
XX PT phosphodiesterase and glycosidase activity and useful in bio:engineering,
XX PT enzymatic and drug screening applications.
XX PS Example 1; Page 14; 35pp; English.
XX CC This sequence represents a specific example of a polypeptide, which forms
XX CC a beta sheet in an aqueous environment and has the general formula
XX CC JnAm1XuAm2Jn or Ac-JnAm1XuAm2Jn-NH2, in which A = D- or L-alanine amino
XX CC acid; m1, m2 = 0-40; m1+m2 = 10-40; J = a charged amino acid; n = 1 or 2;
XX CC X = any amino acid except proline; u = 0 or 1; J1 lysine, hydroxylysine,
XX CC arginine, histidine, aspartic acid, glutamic acid or gamma-
XX CC carboxyglutamic acid; Ac = an acylation modification to the amino
XX CC terminus; NH2 = an amidation modification to the carboxylic acid terminus
XX CC ; and X1 = cysteine, threonine, tyrosine or serine. The peptides have
XX CC phosphodiesterase activity and readily hydrolyse sugar phosphates. They also
XX CC have glycosidase activity and readily hydrolyse nucleic acids. Further
XX CC they can cause amine-catalysed decarboxylation (e.g. decarboxylation of
XX CC oxalacetate), and they can hydrolyse phospholipids. They can be used for
XX CC bioengineering, enzymatic and drug screening applications. In particular
XX CC they can be used for screening drugs for the prevention or treatment of
XX CC neurodegenerative diseases, e.g. Alzheimer's disease, which result in
XX CC amyloid protein deposition. They can also be used to screen for drugs
XX CC which inhibit or disrupt the beta sheet
XX SQ Sequence 16 AA;
XX Query Match 100.0%; Score 44; DB 2; Length 16;
XX Best Local Similarity 84.6%; Pred. No. 2.1;
XX Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX QY 1 AAXAAAAAAAXAA 13
XX DB 2 AAAAAAAAXAAAA 14
XX RESULT 16
XX AA33226
XX ID AA332226 standard; peptide; 16 AA.
XX XX AA332226;
XX AC AA332226;
XX DT 12-FEB-1998 (first entry)
XX XX

```

```

DE XX Beta-sheet forming peptide #2.
XX KW Beta sheet; circular dichroism spectroscopy; glycosidase activity;
XX KW phosphodiesterase activity; drug screening; hydrolysis;
XX KW neurodegenerative disease; Alzheimer's disease; amyloid deposition.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Modified-site 1
XX FT /note= "N-terminal acetyl"
XX FT Modified-site 16
XX FT /note= "C-terminal amide"
XX PN W09637212-A1.
XX XX 28-NOV-1996.
XX PD 23-MAY-1996; 96WO-US007564.
XX PF 26-MAY-1995; 95US-00452043.
XX PR (TORR-) TORREY PINES INST MOLECULAR STUDIES.
XX PA Blondelle SE, Perez-Paya E, Forood B, Houghten RA;
XX PI WPI; 1997-033944/03.
XX DR Polypeptide(s) which form stable beta sheets in aq. environment - having
XX PT phosphodiesterase and glycosidase activity and useful in bio:engineering,
XX PT enzymatic and drug screening applications.
XX PS Example 1; Page 15; 35pp; English.
XX CC This sequence represents a specific example of a polypeptide, which forms
XX CC a beta sheet in an aqueous environment and has the general formula
XX CC JnAm1XuAm2Jn or Ac-JnAm1XuAm2Jn-NH2, in which A = D- or L-alanine amino
XX CC acid; m1, m2 = 0-40; m1+m2 = 10-40; J = a charged amino acid; n = 1 or 2;
XX CC X = any amino acid except proline; u = 0 or 1; J1 lysine, hydroxylysine,
XX CC arginine, histidine, aspartic acid, glutamic acid or gamma-
XX CC carboxyglutamic acid; Ac = an acylation modification to the amino
XX CC terminus; NH2 = an amidation modification to the carboxylic acid terminus
XX CC ; and X1 = cysteine, threonine, tyrosine or serine. The peptides have
XX CC phosphodiesterase activity and readily hydrolyse sugar phosphates. They also
XX CC have glycosidase activity and readily hydrolyse nucleic acids. Further
XX CC they can cause amine-catalysed decarboxylation (e.g. decarboxylation of
XX CC oxalacetate), and they can hydrolyse phospholipids. They can be used for
XX CC bioengineering, enzymatic and drug screening applications. In particular
XX CC they can be used for screening drugs for the prevention or treatment of
XX CC neurodegenerative diseases, e.g. Alzheimer's disease, which result in
XX CC amyloid protein deposition. They can also be used to screen for drugs
XX CC which inhibit or disrupt the beta sheet
XX SQ Sequence 16 AA;
XX Query Match 100.0%; Score 44; DB 2; Length 16;
XX Best Local Similarity 84.6%; Pred. No. 2.1;
XX Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX QY 1 AAXAAAAAAAXAA 13
XX DB 3 AAAAAAAAXAAAA 15
XX RESULT 17
XX ADV78643
XX ID ADV78643 standard; peptide; 16 AA.
XX XX ADV78643;
XX AC ADV78643;
XX DT 24-MAR-2005 (first entry)
XX XX Cell attachment-related auxiliary peptide - SEQ ID 107.
XX DE

```

XX cell attachment; cell culture; transplantation; vaccine; antibiotic.  
 XX Synthetic.  
 XX JP2005002106-A.  
 XX 06-JAN-2005.  
 XX 20-MAY-2004; 2004JP-00150510.  
 XX 21-MAY-2003; 2003JP-00144125.  
 XX (SANN ) SANYO CHEM IND LTD.  
 XX WPI; 2005-053150/06.  
 XX Cell attachable polypeptide useful for culturing cell, comprises  
 PT auxiliary amino acid sequence having specific amount of Gly and Ala,  
 PT chemically bonded to cell attachable minimal amino acid sequence,  
 PT alternatively.  
 XX Disclosure; SEQ ID NO 107; 20pp; Japanese.  
 XX The invention relates to a novel cell attachable polypeptide comprising  
 CC an auxiliary amino acid sequence (Y) chemically bonded to a cell  
 CC attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino  
 CC acids and the sum total content of Gly and Ala in (Y) is 42-100% with  
 CC respect to the total number of amino acids in (Y). The polypeptide of the  
 CC invention may be utilized during culturing of cells for use in various  
 CC experiments, including toxicity studies, gene transfer and production of  
 CC useful substances such as cytokines, thrombolytic agents, vaccines,  
 CC hormones and antibiotics. Furthermore, the polypeptide may be useful for  
 CC culturing cells for transplantation, particularly, transplantation of  
 CC skin, cranial bones, muscles, skin tissue, bone marrow and retina. The  
 CC polypeptide exhibits extremely potent cell adhesion properties and can  
 CC attach cells, such as NHDF and VERO cells, more efficiently onto a base  
 CC material, relative to conventional methods. The polypeptide also exhibits  
 CC high heat stability, thus sterilization in autoclaves can be performed  
 CC conveniently. The current sequence is that of a cell attachment-related  
 CC auxiliary peptide of the invention.  
 XX  
 SQ Sequence 16 AA;  
 Query Match 100.0%; Score 44; DB 9; Length 16;  
 Best Local Similarity 84.6%; Pred. No. 2.1;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAXAAAAAXAA 13  
 |||||  
 Db 1 AAAAAAXAA 13  
 |||||  
 RESULT 18  
 AAW32227  
 ID AAW32227 standard; peptide; 17 AA.  
 XX  
 AC AAW32227;  
 XX  
 DT 12-FEB-1998 (first entry)  
 XX  
 DE Beta-sheet forming peptide #3.  
 XX  
 KW Beta sheet; circular dichroism spectroscopy; glycosidase activity;  
 KW phosphodiesterase activity; drug screening; hydrolysis;  
 KW neurodegenerative disease; Alzheimer's disease; amyloid deposition.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Modified-site 1  
 FT Modified-site /note= "N-terminal acetyl"  
 FT Modified-site 17

FT  
 XX  
 PN  
 XX  
 PD  
 XX  
 XX 28-NOV-1996.  
 XX 23-MAY-1996; 96WO-US007564.  
 XX 26-MAY-1995; 95US-00452043.  
 XX (TORR-) TORREY PINES INST MOLECULAR STUDIES.  
 XX  
 PI Blondelle SE, Perez-Paya E, Forood B, Houghten RA;  
 XX  
 XX WPI; 1997-033944/03.  
 XX  
 XX Polypeptide(s) which form stable beta sheets in aq. environment - having  
 PT phosphodiesterase and glycosidase activity and useful in bioengineering,  
 PT enzymatic and drug screening applications.  
 XX  
 XX Example 2; Page 16; 35pp; English.  
 XX  
 CC This sequence represents a specific example of a polypeptide, which forms  
 CC a beta sheet in an aqueous environment and has the general formula  
 CC JnAmXuAm2Jn or Ac-JinAmXuAm2Jn-NH2, in which A = D- or L-alanine amino  
 CC acid; m1, m2 = 0-40; m1+m2 = 10-40; J = a charged amino acid; n = 1 or 2;  
 CC X = any amino acid except proline; u = 0 or 1; J1 lysine, hydroxylysine,  
 CC arginine, histidine, aspartic acid, glutamic acid or gamma-  
 CC carboxyglutamic acid; Ac = an acylation modification to the amino  
 CC terminus; NH2 = an amidation modification to the carboxylic acid terminus  
 CC ; and X1 = cysteine, threonine, tyrosine or serine. The peptides have  
 CC phosphodiesterase activity and readily hydrolyse nucleic acids. They also  
 CC have glycosidase activity and readily hydrolyse sugar phosphates. Further  
 CC they can cause amine-catalysed decarboxylation (e.g. decarboxylation of  
 CC oxalacetate), and they can hydrolyse phospholipids. They can be used for  
 CC bioengineering, enzymatic and drug screening applications. In particular  
 CC they can be used for screening drugs for the prevention or treatment of  
 CC neurodegenerative diseases, e.g. Alzheimer's disease, which result in  
 CC amyloid protein deposition. They can also be used to screen for drugs  
 CC which inhibit or disrupt the beta sheet  
 XX  
 SQ Sequence 17 AA;  
 Query Match 100.0%; Score 44; DB 2; Length 17;  
 Best Local Similarity 84.6%; Pred. No. 2.2;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 AAXAAAAAXAA 13  
 |||||  
 Db 3 AAAAAAXAA 15  
 |||||  
 RESULT 19  
 AAO30390  
 ID AAO30390 standard; peptide; 18 AA.  
 XX  
 AC AAO30390;  
 XX  
 DT 22-SEP-2003 (first entry)  
 XX  
 DE Human ARX (GCG)10+7 mutant cDNA fragment encoded peptide.  
 XX  
 KW Homeobox gene; ARX; ARX-related disorder; X-linked myoclonic epilepsy;  
 KW infantile spasm; mental retardation; Partington syndrome; diagnosis;  
 KW dystonia; human; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX WO2003045989-A1.  
 XX  
 PD 05-JUN-2003.  
 XX

/note= "C-terminal amide"

PF 26-NOV-2002; 2002WO-AU001599.  
XX  
PR 26-NOV-2001; 2001AU-00009095.  
XX  
XX (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.  
XX  
XX Gecz J, Stromme P;  
XX  
XX WPI; 2003-505184/47.  
DR N-PSDB; AAL61167.  
XX  
XX New ARX gene, useful for diagnosing ARX-related disorders e.g., mild  
PT mental retardation, infantile spasms or dystonia.  
XX  
XX Disclosure; Fig 2a; 74pp; English.  
PS  
XX The invention relates to human orthologue of Aristalless homeobox gene,  
CC ARX. ARX gene is associated with infantile spasms (IS), non-specific X-  
CC linked mental retardation, X-linked myoclonic epilepsy and Partington  
CC syndrome. ARX gene is useful for diagnosing ARX-related disorders e.g.,  
CC mild mental retardation, infantile spasms or dystonia. The present  
CC sequence is human ARX (GCG)10+7 mutant cDNA fragment encoded peptide  
XX  
XX Sequence 18 AA;  
SQ  
Query Match 100.0%; Score 44; DB 7; Length 18;  
Best Local Similarity 84.6%; Pred. No. 2.3;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAXAAAAAAAXAA 13  
DB 1 AAAAAAAAAXAA 13  
RESULT 20  
ADH58940  
ID ADH58940 standard; peptide; 18 AA.  
XX  
AC ADH58940;  
XX  
DT 25-MAR-2004 (first entry)  
XX  
DE Silk protein related peptide #SEQ ID 9.  
XX  
XX Silk; genetically modified; silk-like protein; polymer; silkworm fibroin;  
KW elastin; fibronectin; textile; food; industry; cosmetics; PET-SLP(n);  
KW PET-SLP(2; 4; 6); SLP6.  
XX  
XX Synthetic.  
OS  
XX  
XX WO2003100065-A1.  
PN  
XX  
XX 04-DEC-2003.  
PD  
XX  
XX 23-MAY-2002; 2002WO-JP005010.  
PF  
XX  
XX 23-MAY-2002; 2002WO-JP005010.  
PR  
XX  
XX (UNNI-) UNIV JAPAN TOKYO AGR.  
FA  
XX  
XX Asakura T;  
PI  
XX  
XX WPI; 2004-108256/11.  
DR  
XX  
XX Mass production of silk protein and silk-like protein having imparted  
PT function by employing genetically-modified Escherichia coli, useful in  
PT textiles, food industry and cosmetics.  
XX  
XX Disclosure; SEQ ID NO 9; 56pp; Japanese.  
PS  
XX  
XX The invention relates to a method for producing silk or genetically-  
CC modified silk-like protein, comprising designing a silk or silk-like  
CC polymer containing e.g. domesticated silkworm fibroin and synthesizing  
CC

CC the minimum unit of the silk or the thus designed polymer for integration  
CC into expression vectors and transfer into Escherichia coli prior to  
CC growing in a medium. Also disclosed is a method for producing silk or  
CC genetically-modified silk-like protein, comprising designing a silk or  
CC silk-like polymer containing at least one protein selected from  
CC domesticated silkworm fibroin, wild silkworm fibroin, elastin and  
CC fibronectin and essentially comprising the domesticated or wild silkworm  
CC fibroin. The produced silk and silk-like protein are useful in textiles,  
CC food industry and cosmetics. Using this method, expression efficiency is  
CC greatly improved, thus providing silk-like protein with designed  
CC functionalities easily in large quantities. The SLP gene, PUC-link and  
CC SLP(n) were constructed for producing expression vector PET-SLP(n) with  
CC expression of PET-SLP(2,4,6). The target protein SLP6 was obtained after  
CC cleavage of the tag sequence with use of cyanogens bromide. The sequences  
CC given in records for ADH58932-ADH58971 represent silk-like protein  
CC related polypeptides and polynucleotides, and primers used in the  
CC creation of the vector of the invention.  
XX  
XX Sequence 18 AA;  
SQ  
Query Match 100.0%; Score 44; DB 8; Length 18;  
Best Local Similarity 84.6%; Pred. No. 2.3;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAXAAAAAAAXAA 13  
DB 1 AAAAAAAAAXAA 13  
RESULT 21  
AEF40057  
ID AEF40057 standard; peptide; 18 AA.  
XX  
AC AEF40057;  
XX  
DT 23-MAR-2006 (first entry)  
XX  
DE Polyalanine peptide.  
XX  
KW Protein production; silk; food; cosmetics; polyalanine.  
XX  
XX Unidentified.  
OS  
XX  
XX US2006019348-A1.  
PN  
XX  
XX 26-JAN-2006.  
PD  
XX  
XX 05-AUG-2005; 2005US-00197315.  
PF  
XX  
XX 23-MAY-2002; 2002WO-JP005010.  
PR  
XX  
XX 22-NOV-2004; 2004US-00515264.  
XX  
XX (ASAK/) ASAKURA T.  
PA  
XX  
XX Asakura T;  
PI  
XX  
XX WPI; 2006-109210/11.  
DR  
XX  
XX Producing silk or silk-like protein comprises integrating a silk like  
PT polymer comprising the minimum unit into an expression vector selected  
PT from expression vectors containing T7 promoter.  
XX  
XX Disclosure; SEQ ID NO 9; 25pp; English.  
PS  
XX  
XX The present invention provides a method for mass production of silk  
CC protein and gene recombinant silk-like protein (SLP). The method involves  
CC integrating a silk like polymer selected from e.g. domestic silkworm  
CC fibroin, wild silkworm fibroin, elastin and fibronectin of the minimum  
CC unit into an expression vector selected from expression vectors  
CC containing T7 promoter. The method is useful for mass producing silk or  
CC silk-like proteins with improved cellular adhesiveness, elasticity or  
CC hardness and is also useful in clothing, food and cosmetic industries.  
CC The present sequence is a polyalanine peptide which is similar to

ID ADV78587 standard; peptide; 19 AA.



PN JP2005170810-A.  
XX  
PD 30-JUN-2005.  
XX  
XX 09-DEC-2003; 2003JP-00409854.  
XX  
XX 09-DEC-2003; 2003JP-00409854.  
PR  
XX (SANN ) SANYO CHEM IND LTD.  
XX  
XX Kurokawa Y;  
PI  
XX WPI; 2005-514832/53.  
DR  
XX Novel polypeptide having cell adhesion property, useful for culturing  
PT cells and producing bioactive substances.  
PT  
XX Disclosure; SEQ ID NO 63; 26pp; Japanese.  
XX  
XX The invention relates to a novel polypeptide having cell adhesion  
CC property. The polypeptide comprises a minimum amino acid sequence and an  
CC auxiliary amino acid sequence capable of expressing a cell adhesion  
CC property and having a molecular weight of 1500-15000. The invention  
CC further comprises a cell adhesion property polypeptide base material. The  
CC cell adhesion property polypeptide containing base material is useful for  
CC producing a cell. The cell adhesion property polypeptide is useful for  
CC culturing and analyzing cells, screening assays, and producing bioactive  
CC substances such as cytokines, hormones, antibiotics, etc. The cell  
CC adhesion property polypeptide has excellent cell adhesion property, heat  
CC stability and has low immunogenicity. This sequence represents a human  
CC peptide used as the auxiliary amino acid sequence used in the cell  
CC adhesion property polypeptide of the invention.  
XX  
SQ Sequence 19 AA;  
  
Query Match 100.0%; Score 44; DB 9; Length 19;  
Best Local Similarity 84.6%; Pred. No. 2.4;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAXAAAAAAXAA 13  
||| ||||| ||  
DB 1 AAAAAAAXAAAAA 13  
  
RESULT 25  
AAB69489  
ID AAB69489 standard; peptide; 20 AA.  
XX  
AC AAB69489;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE Synthetic HAV peptide, SEQ ID NO: 89.  
XX  
XX Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;  
KW antigen; major structural capsid polypeptide; HAV antibody detection.  
XX  
XX Hepatitis A virus.  
OS Synthetic.  
OS  
XX WO200105824-A2.  
XX  
XX 25-JAN-2001.  
XX  
XX 14-JUL-2000; 2000WO-US019267.  
XX  
XX 15-JUL-1999; 99US-0144412P.  
XX  
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
XX Fields HA, Khudyakov YE;  
PI  
XX WPI; 2001-112681/12.  
DR

XX Synthetic peptides used as antigen sources for enzyme immunoassays  
PT detecting anti-hepatitis A virus and as vaccines.  
XX  
XX Disclosure; Page 127; 130pp; English.  
XX  
XX The present sequence is one of a number of synthetic peptides which are  
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides  
CC comprise antigenic epitopes of the major structural capsid polypeptides  
CC or non-structural polypeptides of HAV with one or more glutamine  
CC molecules at the carboxy end of the peptide. The peptides are used to  
CC detect the presence of antibodies against HAV in mammalian serum, to  
CC detect the presence of HAV in a human or animal through the binding of  
CC the peptide to an antibody, to detect acute phase infection by detecting  
CC IGM antibodies in mammalian serum and detecting convalescence in a  
CC mammal. The peptides are used to detect or quantify HAV antibodies in  
CC samples in clinical or research-based assays using immunoblotting,  
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,  
CC tracking of radioactive or bioluminescent markers, chromatography or  
CC electrophoresis. The peptides are used to induce an immune response to  
CC HAV when administered to a human or animal. Glutamine at the carboxy end  
CC of the peptides enhances the IGM antibody reactivity  
XX  
SQ Sequence 20 AA;  
  
Query Match 100.0%; Score 44; DB 4; Length 20;  
Best Local Similarity 84.6%; Pred. No. 2.5;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 AAXAAAAAAXAA 13  
||| ||||| ||  
DB 1 AAAAAAAXAAAAA 13  
  
RESULT 26  
ADH47641  
ID ADH47641 standard; peptide; 20 AA.  
XX  
XX ADH47641;  
AC  
XX 25-MAR-2004 (first entry)  
DT  
XX Human lung tumour clone peptide, SEQ ID No 2122.  
DE  
XX lung cancer; tumour; cytostatic; immunostimulant; vaccine; gene therapy;  
KW human; clone.  
XX  
XX Homo sapiens.  
OS  
XX WO2003037267-A2.  
PN  
XX 08-MAY-2003.  
PD  
XX 28-OCT-2002; 2002WO-US034777.  
PF  
XX 29-OCT-2001; 2001US-00017754.  
XX  
PR 28-MAR-2002; 2002US-00013872.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Henderson RA, Wang T, Watanabe Y, Kalos MD, Sleath PR;  
PI Johnson JC, Retter MW, Durham M, Carter D, Fanger GR, vedvick TS;  
PI Bangur CS, McNabb A;  
XX  
XX WPI; 2003-468346/44.  
DR  
XX New polypeptides and encoding polynucleotides, useful for diagnosing,  
PT preventing and/or treating lung cancer.  
PT  
XX Disclosure; SEQ ID NO 2122; 258pp; English.  
XX  
XX The invention relates to novel compositions and methods for the therapy  
CC and diagnosis of cancer, particularly lung cancer. The compositions  
CC

CC comprise one or more lung tumour polypeptides, immunogenic portions  
 CC thereof, polynucleotides that encode such polypeptides, antigen presenting  
 CC cells that express such polypeptides, and T cells that are specific for  
 CC cells expressing such polypeptides. The novel compositions have  
 CC cytostatic and immunostimulant activity. The lung tumour antigens can be  
 CC used in the creation of a vaccine. The polynucleotides that encode the  
 CC lung tumour polypeptides can be used in gene therapy to help in the  
 CC treatment of lung tumours. This sequence represents a human lung tumour  
 CC clone polypeptide of the invention. This sequence was not shown in the  
 CC specification. It has been taken from a World Intellectual Property  
 CC Organization CD ROM supplied with the specification.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 44; DB 7; Length 20;  
 Best Local Similarity 84.6%; Pred. No. 2.5;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAXAAXAA 13  
 DB 5 AAAAAAAAAAAAAA 17

RESULT 27

ADH58955  
 ID ADH58955 standard; peptide; 20 AA.

XX AC

ADH58955;

DT 25-MAR-2004 (first entry)

XX Silk protein related peptide #SEQ ID 24.

DE Silk; genetically modified; silk-like protein; polymer; silkworm fibroin;  
 KW elastin; fibronectin; textile; food; industry; cosmetics; PET-SLP(n);  
 KW PET-SLP(2; 4; 6); SLP6.

XX Synthetic.

XX WO2003100065-A1.

XX PD 04-DEC-2003.

XX 23-MAY-2002; 2002WO-JP005010.

XX 23-MAY-2002; 2002WO-JP005010.

XX (UYN1-) UNIV JAPAN TOKYO AGR.

XX Asakura T;

XX WPI; 2004-108256/11.

XX Mass production of silk protein and silk-like protein having imparted  
 PT function by employing genetically-modified *Escherichia coli*, useful in  
 PT textiles, food industry and cosmetics.

XX Example 2; SEQ ID NO 24; 56pp; Japanese.

XX The invention relates to a method for producing silk or genetically-  
 CC modified silk-like protein, comprising designing a silk or silk-like  
 CC polymer containing e.g. domesticated silkworm fibroin and synthesizing  
 CC the minimum unit of the silk or the thus designed polymer for integration  
 CC into expression vectors and transfer into *Escherichia coli* prior to  
 CC growing in a medium. Also disclosed is a method for producing a silk or  
 CC genetically-modified silk-like protein, comprising designing a silk or  
 CC silk-like polymer containing at least one protein selected from  
 CC domesticated silkworm fibroin, wild silkworm fibroin, elastin and  
 CC fibronectin and essentially comprising the domesticated or wild silkworm  
 CC fibroin. The produced silk and silk-like protein are useful in textiles,  
 CC food industry and cosmetics. Using this method, expression efficiency is  
 CC greatly improved, thus providing silk-like protein with designed  
 CC functionalities easily in large quantities. The SLP gene, PUC-link and

CC SLP(n) were constructed for producing expression vector PET-SLP(n) with  
 CC expression of PET-SLP(2,4,6). The target protein SLP6 was obtained after  
 CC cleavage of the tag sequence with use of cyanogens bromide. The sequences  
 CC given in records for ADH58932-ADH58971 represent silk-like protein  
 CC related polypeptides and polynucleotides, and primers used in the  
 CC creation of the vector of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 44; DB 8; Length 20;  
 Best Local Similarity 84.6%; Pred. No. 2.5;  
 Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAXAAXAAXAA 13  
 DB 3 AAAAAAAAAAAAAA 15

RESULT 28

ADJ21560

ID ADJ21560 standard; peptide; 20 AA.

XX AC

ADJ21560;

DT 04-NOV-2004 (first entry)

XX Human lung cancer-related L984P peptide fragment - SEQ ID 2122.

XX lung tumour; cytostatic; lung cancer; human; L984P.

XX Homo sapiens.

XX US2003211510-A1.

XX PD 13-NOV-2003.

XX 28-OCT-2002; 2002US-00283017.

XX 30-JUN-1999; 99US-00346492.

XX 15-OCT-1999; 99US-00419356.

XX 17-DEC-1999; 99US-00468867.

XX 30-DEC-1999; 99US-00476300.

XX 06-MAR-2000; 2000US-00519642.

XX 22-MAR-2000; 2000US-00533077.

XX 10-APR-2000; 2000US-00546259.

XX 27-APR-2000; 2000US-00560406.

XX 05-JUN-2000; 2000US-00589184.

XX 11-JUL-2000; 2000US-00614124.

XX 29-AUG-2000; 2000US-00651563.

XX 08-SEP-2000; 2000US-00658824.

XX 26-SEP-2000; 2000US-00671325.

XX 06-OCT-2000; 2000US-00677419.

XX 30-OCT-2000; 2000US-00702705.

XX 13-DEC-2000; 2000US-00736457.

XX 03-MAY-2001; 2001US-00849626.

XX 10-JUL-2001; 2001US-00902941.

XX 29-OCT-2001; 2001US-00017754.

XX 28-MAR-2002; 2002US-00113872.

XX (CORI-) CORIXA CORP.

XX Henderson RA, Wang T, Watanabe Y, Kalos MD, Sleath PR;  
 PI Johnson JC, Retter MW, Durham M, Carter D, Fanger GR, Vedvick TS;  
 PI Bangur CS, McNabb A;  
 DR WPI; 2004-167010/16.  
 XX Novel polynucleotide encoding lung tumor polypeptides, useful for  
 PT diagnosing, preventing and treating cancer e.g. lung cancer.  
 XX Example 39; SEQ ID NO 2122; 99pp; English.  
 CC The invention relates to a novel isolated polynucleotide comprising a

CC sequence chosen from any one of 40 lung tumour polypeptides or its  
CC complements, fragments or degenerate variants. The method of the  
CC invention has cytostatic applications and may be useful for detecting and  
CC treating lung cancer in a patient, as well as for inhibiting the  
CC development of lung cancer in a patient via incubating CD4+ and/or CD8+ T  
CC cells isolated from a patient with at least one component chosen from a  
CC polypeptide, polynucleotide or antigen presenting cell (APC) of the  
CC invention and administering an effective amount of the proliferated T  
CC cells to the patient. The current sequence is that of the human lung  
CC cancer-related peptide of the invention. The current sequence is not  
CC shown in the specification per se but is available on the USPTO web-site  
CC <http://seqdata.uspto.gov/sequence.html?DocID=20030211510>.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 44; DB 8; Length 20;  
Best Local Similarity 84.6%; Pred. No. 2.5;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAXAA 13  
||| ||||| ||  
Db 5 AAAAAAAXAA 17

RESULT 29

ADV78588  
ID ADV78588 standard; peptide; 20 AA.

XX AC ADV78588;

XX DT 24-MAR-2005 (first entry)

XX DE Cell attachment-related auxillary peptide - SEQ ID 52.

XX KW cell attachment; cell culture; transplantation; vaccine; antibiotic.

XX OS Synthetic.

XX PN JP2005002106-A.

XX PD 06-JAN-2005.

XX PF 20-MAY-2004; 2004JP-00150510.

XX PR 21-MAY-2003; 2003JP-00144125.

XX PA (SANN ) SANYO CHEM IND LTD.

XX PS WPI; 2005-053150/06.

XX PT Cell attachable polypeptide useful for culturing cell, comprises  
XX an auxillary amino acid sequence having specific amount of Gly and Ala,  
XX chemically bonded to cell attachable minimal amino acid sequence,  
XX alternatively.

XX PS Disclosure; SEQ ID NO 52; 20pp; Japanese.

XX CC The invention relates to a novel cell attachable polypeptide comprising  
XX an auxillary amino acid sequence (Y) chemically bonded to a cell  
XX attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino  
XX acids and the sum total content of Gly and Ala in (Y) is 42-100% with  
XX respect to the total number of amino acids in (Y). The polypeptide of tHe  
XX invention may be utilized during culturing of cells for use in various  
XX experiments, including toxicity studies, gene transfer and production of  
XX useful substances such as cytokines, thrombolytic agents, vaccines,  
XX hormones and antibiotics. Furthermore, the polypeptide may be useful for  
XX culturing cells for transplantation, particularly, transplantation of  
XX skin, cranial bones, muscles, skin tissue, bone marrow and retina. The  
XX polypeptide exhibits extremely potent cell adhesion properties and can  
XX attach cells, such as NHDF and VERO cells, more efficiently onto a base  
XX material, relative to conventional methods. The polypeptide also exhibits  
XX high heat stability, thus sterilization in autoclaves can be performed  
XX conveniently. The current sequence is that of a cell attachment-related

CC auxillary peptide of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 44; DB 9; Length 20;  
Best Local Similarity 84.6%; Pred. No. 2.5;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAXAA 13  
||| ||||| ||  
Db 4 AAAAAAAXAA 16

RESULT 30

AEF40072  
ID AEF40072 standard; peptide; 20 AA.

XX AC AEF40072;

XX DT 23-MAR-2006 (first entry)

XX DE Domestic silkworm silk-like protein with polyalanine peptide SEQ ID: 24.

XX KW protein production; silk; food; cosmetics; silk-like protein;  
XX polyalanine; fusion protein.

XX OS Bombyx mori.

XX OS Unidentified.

XX PN US2006019348-A1.

XX PD 26-JAN-2006.

XX PF 05-AUG-2005; 2005US-00197315.

XX PR 23-MAY-2002; 2002WO-JP005010.

XX PR 22-NOV-2004; 2004US-00515264.

XX PA (ASAK/) ASAKURA T.

XX PI Asakura T;

XX DR WPI; 2006-109210/11.

XX DR N-PSDB; AEF40068.

XX PT Producing silk or silk-like protein comprises integrating a silk like  
XX polymer comprising the minimum unit into an expression vector selected  
XX from expression vectors containing T7 promoter.

XX PS Example 2; SEQ ID NO 24; 25pp; English.

XX CC The present invention provides a method for mass production of silk  
XX protein and gene recombinant silk-like protein (SLP). The method involves  
XX integrating a silk like polymer selected from e.g. domestic silkworm  
XX fibroin, wild silkworm fibroin, elastin and fibronectin of the minimum  
XX unit into an expression vector selected from expression vectors  
XX containing T7 promoter. The method is useful for mass producing silk or  
XX silk-like proteins with improved cellular adhesiveness, elasticity or  
XX hardness and is also useful in clothing, food and cosmetic industries.  
XX The present sequence is a silk-like protein with polyalanine (SLPA)  
XX peptide.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 44; DB 10; Length 20;  
Best Local Similarity 84.6%; Pred. No. 2.5;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAXAAAAAAXAA 13  
||| ||||| ||  
Db 3 AAAAAAAXAA 15

```
RESULT 31
AAY25436
ID AAY25436 standard; protein; 21 AA.
XX
AC AAY25436;
XX
DT 09-SEP-1999 (first entry)
XX
DE Synthetic antifreeze protein APWF8R insert fragment B2.
XX
KW Antifreeze peptide; AFP; wfsr; APWF8R; winter flourider; ice formation;
KW water freezing temperature; antifreeze agent.
XX
OS Synthetic.
XX
PN US925540-A.
XX
PD 20-JUL-1999.
XX
PF 23-DEC-1991; 91US-00814220.
XX
PR 25-SEP-1990; 90US-00598437.
XX
PA (VIRG ) VIRGINIA TECH INTELLECTUAL PROPERTIES.
XX
PI Szumanski MBW, Toth TE, Caceci T;
XX
DR WPI; 1999-418272/35.
XX
DR N-PSDB; AAX88104, AAX88105.
XX
PT Synthetic antifreeze peptide and the gene coding its production.
XX
PS Disclosure; Fig 8C; 29pp; English.
XX
CC This invention describes the production of a novel synthetic DNA (1),
CC wfsr, encoding an antifreeze polypeptide, APWF8R modelled on the
CC antifreeze peptide (AFP) found in winter flourider. The polypeptides
CC encoded by (1) lower the freezing temperature of water and prevent ice
CC formation, therefore are useful as antifreeze agents. The use of
CC synthetic peptides is less toxic than the use of chemicals (prior art).
CC Naturally occurring antifreeze polypeptides are not available in large
CC quantities, synthetic antifreeze polypeptides are more readily available.
CC The antifreeze polypeptides are produced by recombinant means, a process
CC which is cheaper and less time consuming compared to direct chemical
CC synthesis of synthetic peptides (prior art)
XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 44; DB 2; Length 21;
Best Local Similarity 84.6%; Pred. NO. 2.6;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAXAAAAAAAXAA 13
DB 8 AATAAAAAATAA 20
RESULT 32
AAB69490
ID AAB69490 standard; peptide; 21 AA.
XX
AC AAB69490;
XX
DT 20-APR-2001 (first entry)
XX
DE Synthetic HAV peptide, SEQ ID NO: 90.
XX
KW Hepatitis A virus; HAV; immunogen; immunostimulant; virucide; vaccine;
KW antigen; major structural capsid polypeptide; HAV antibody detection.
XX
OS Hepatitis A virus.
OS Synthetic.
XX
PN WO200105824-A2.
XX
PD 25-JAN-2001.
XX
PF 14-JUL-2000; 2000WO-US019267.
XX
PR 15-JUL-1999; 99US-0144412P.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Fields HA, Khudyakov YE;
XX
DR WPI; 2001-112681/12.
XX
PT Synthetic peptides used as antigen sources for enzyme immunoassays
PT detecting anti-hepatitis A virus and as vaccines.
XX
PS Disclosure; Page 127; 130pp; English.
XX
CC The present sequence is one of a number of synthetic peptides which are
CC immunoreactive with hepatitis A virus (HAV) antibodies. The peptides
CC comprise antigenic epitopes of the major structural capsid polypeptides
CC or non-structural polypeptides of HAV with one or more glutamine
CC molecules at the carboxy end of the peptide. The peptides are used to
CC detect the presence of antibodies against HAV in mammalian serum, to
CC detect the presence of HAV in a human or animal through the binding of
CC the peptide to an antibody, to detect acute phase infection by detecting
CC IGM antibodies in mammalian serum and detecting convalescence in a
CC mammal. The peptides are used to detect or quantify HAV antibodies in
CC samples in clinical or research-based assays using immunoblotting,
CC fluorescence in situ hybridisation analysis, gel-mobility shift assays,
CC tracking of radioactive or bioluminescent markers, chromatography or
CC electrophoresis. The peptides are used to induce an immune response to
CC HAV when administered to a human or animal. Glutamine at the carboxy end
CC of the peptides enhances the IGM antibody reactivity
XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 44; DB 4; Length 21;
Best Local Similarity 84.6%; Pred. NO. 2.6;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAXAAAAAAAXAA 13
DB 1 AAAAAAAAXAA 13
RESULT 33
ADJ81212
ID ADJ81212 standard; peptide; 21 AA.
XX
AC ADJ81212;
XX
DT 06-MAY-2004 (first entry)
XX
DE Self-assembling polymer peptide #67.
XX
KW miniblock polymer; optical response; mid infrared wavelength range;
KW long range ordered fluid; liquid crystals; nanolithography;
KW IR-sensitive device; IR sensor; IR filter; night telescope;
KW thermosensitive detector; non-linear chromophore; biomaterial;
KW tissue engineering scaffold; ferroelectric material; artificial muscle;
KW switching device.
XX
OS Synthetic.
XX
PN WO2003056297-A2.
XX
PD 10-JUL-2003.
XX
PF 02-OCT-2002; 2002WO-US031375.
XX
PR 02-OCT-2001; 2001US-0326743P.
```



DE Cell attachment-related auxillary peptide - SEQ ID 53.  
KW cell attachment; cell culture; transplantation; vaccine; antibiotic.  
XX Synthetic.  
OS  
XX JP2005002106-A.  
PN  
XX  
XX 06-JAN-2005.  
PD  
XX 20-MAY-2004; 2004JP-00150510.  
XX  
XX 21-MAY-2003; 2003JP-00144125.  
XX  
XX (SANN ) SANYO CHEM IND LTD.  
PA  
XX WPI; 2005-053150/06.  
XX  
XX Cell attachable polypeptide useful for culturing cell, comprises  
PT auxillary amino acid sequence having specific amount of Gly and Ala,  
PT chemically bonded to cell attachable minimal amino acid sequence,  
PT alternatively.  
XX  
XX Disclosure; SEQ ID NO 53; 20pp; Japanese.  
XX  
XX The invention relates to a novel cell attachable polypeptide comprising  
CC an auxillary amino acid sequence (Y) chemically bonded to a cell  
CC attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino  
CC acids and the sum total content of Gly and Ala in (Y) is 42-100% with  
CC respect to the total number of amino acids in (Y). The polypeptide of the  
CC invention may be utilized during culturing of cells for use in various  
CC experiments, including toxicity studies, gene transfer and production of  
CC useful substances such as cytokines, thrombolytic agents, vaccines,  
CC hormones and antibiotics. Furthermore, the polypeptide may be useful for  
CC culturing cells for transplantation, particularly, transplantation of  
CC skin, cranial bones, muscles, skin tissue, bone marrow and retina. The  
CC polypeptide exhibits extremely potent cell adhesion properties and can  
CC attach cells, such as NHDF and VERO cells, more efficiently onto a base  
CC material, relative to conventional methods. The polypeptide also exhibits  
CC high heat stability, thus sterilization in autoclaves can be performed  
CC conveniently. The current sequence is that of a cell attachment-related  
CC auxillary peptide of the invention.  
XX  
SQ Sequence 21 AA;  
Query Match 100.0%; Score 44; DB 9; Length 21;  
Best Local Similarity 84.6%; Pred.No. 2.6;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAXAAAAAAXAA 13  
|| || || || || || || ||  
Db 4 AAAAAAAXAAAAA 16  
RESULT 36  
ADV78590  
ID ADV78590 standard; peptide; 22 AA.  
XX  
AC ADV78590;  
XX  
XX 24-MAR-2005 (first entry)  
DT  
XX  
XX Cell attachment-related auxillary peptide - SEQ ID 54.  
DE  
XX cell attachment; cell culture; transplantation; vaccine; antibiotic.  
KW  
XX Synthetic.  
OS  
XX JP2005002106-A.  
PN  
XX  
XX 06-JAN-2005.  
PD  
XX 20-MAY-2004; 2004JP-00150510.  
XX

XX 21-MAY-2003; 2003JP-00144125.  
PR (SANN ) SANYO CHEM IND LTD.  
XX  
XX WPI; 2005-053150/06.  
XX  
XX Cell attachable polypeptide useful for culturing cell, comprises  
PT auxillary amino acid sequence having specific amount of Gly and Ala,  
PT chemically bonded to cell attachable minimal amino acid sequence,  
PT alternatively.  
XX  
XX Disclosure; SEQ ID NO 54; 20pp; Japanese.  
XX  
XX The invention relates to a novel cell attachable polypeptide comprising  
CC an auxillary amino acid sequence (Y) chemically bonded to a cell  
CC attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino  
CC acids and the sum total content of Gly and Ala in (Y) is 42-100% with  
CC respect to the total number of amino acids in (Y). The polypeptide of the  
CC invention may be utilized during culturing of cells for use in various  
CC experiments, including toxicity studies, gene transfer and production of  
CC useful substances such as cytokines, thrombolytic agents, vaccines,  
CC hormones and antibiotics. Furthermore, the polypeptide may be useful for  
CC culturing cells for transplantation, particularly, transplantation of  
CC skin, cranial bones, muscles, skin tissue, bone marrow and retina. The  
CC polypeptide exhibits extremely potent cell adhesion properties and can  
CC attach cells, such as NHDF and VERO cells, more efficiently onto a base  
CC material, relative to conventional methods. The polypeptide also exhibits  
CC high heat stability, thus sterilization in autoclaves can be performed  
CC conveniently. The current sequence is that of a cell attachment-related  
CC auxillary peptide of the invention.  
XX  
SQ Sequence 22 AA;  
Query Match 100.0%; Score 44; DB 9; Length 22;  
Best Local Similarity 84.6%; Pred.No. 2.8;  
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AAXAAAAAAXAA 13  
|| || || || || || || ||  
Db 4 AAAAAAAXAAAAA 16  
RESULT 37  
ADV78591  
ID ADV78591 standard; peptide; 23 AA.  
XX  
AC ADV78591;  
XX  
XX 24-MAR-2005 (first entry)  
DT  
XX  
XX Cell attachment-related auxillary peptide - SEQ ID 55.  
DE  
XX cell attachment; cell culture; transplantation; vaccine; antibiotic.  
KW  
XX Synthetic.  
OS  
XX JP2005002106-A.  
PN  
XX  
XX 06-JAN-2005.  
PD  
XX  
XX 20-MAY-2004; 2004JP-00150510.  
XX  
XX 21-MAY-2003; 2003JP-00144125.  
XX  
XX (SANN ) SANYO CHEM IND LTD.  
PA  
XX WPI; 2005-053150/06.  
XX  
XX Cell attachable polypeptide useful for culturing cell, comprises  
PT auxillary amino acid sequence having specific amount of Gly and Ala,  
PT chemically bonded to cell attachable minimal amino acid sequence,  
PT alternatively.  
XX

XX PS Disclosure; SEQ ID NO 55; 20pp; Japanese.

XX CC The invention relates to a novel cell attachable polypeptide comprising an auxiliary amino acid sequence (Y) chemically bonded to a cell attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino acids and the sum total content of Gly and Ala in (Y) is 42-100% with respect to the total number of amino acids in (Y). The polypeptide of the invention may be utilized during culturing of cells for use in various experiments, including toxicity studies, gene transfer and production of useful substances such as cytokines, thrombolytic agents, vaccines, hormones and antibiotics. Furthermore, the polypeptide may be useful for culturing cells for transplantation, particularly, transplantation of skin, cranial bones, muscles, skin tissue, bone marrow and retina. The polypeptide exhibits extremely potent cell adhesion properties and can attach cells, such as NHDF and VERO cells, more efficiently onto a base material, relative to conventional methods. The polypeptide also exhibits high heat stability, thus sterilization in autoclaves can be performed conveniently. The current sequence is that of a cell attachment-related auxiliary peptide of the invention.

XX CC

SQ Sequence 23 AA;

Query Match 100.0%; Score 44; DB 9; Length 23;

Best Local Similarity 84.6%; Pred. No. 2.9;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||

Db 4 AAAAAAAAXAA 16

RESULT 38

ADV78592

ID ADV78592 standard; peptide; 24 AA.

XX AC ADV78592;

XX DT 24-MAR-2005 (first entry)

XX DE Cell attachment-related auxiliary peptide - SEQ ID 56.

XX KW cell attachment; cell culture; transplantation; vaccine; antibiotic.

XX OS Synthetic.

XX PN JP2005002106-A.

XX PD 06-JAN-2005.

XX PF 20-MAY-2004; 2004JP-00150510.

XX PR 21-MAY-2003; 2003JP-00144125.

XX PA (SANN) SANYO CHEM IND LTD.

XX DR WPI; 2005-053150/06.

XX CC Cell attachable polypeptide useful for culturing cell, comprises auxiliary amino acid sequence having specific amount of Gly and Ala, chemically bonded to cell attachable minimal amino acid sequence, alternatively.

XX PS Disclosure; SEQ ID NO 56; 20pp; Japanese.

XX CC The invention relates to a novel cell attachable polypeptide comprising an auxiliary amino acid sequence (Y) chemically bonded to a cell attachable minimal amino acid sequence (X), where (Y) contains 1-50 amino acids and the sum total content of Gly and Ala in (Y) is 42-100% with respect to the total number of amino acids in (Y). The polypeptide of the invention may be utilized during culturing of cells for use in various experiments, including toxicity studies, gene transfer and production of useful substances such as cytokines, thrombolytic agents, vaccines,

CC hormones and antibiotics. Furthermore, the polypeptide may be useful for culturing cells for transplantation, particularly, transplantation of skin, cranial bones, muscles, skin tissue, bone marrow and retina. The polypeptide exhibits extremely potent cell adhesion properties and can attach cells, such as NHDF and VERO cells, more efficiently onto a base material, relative to conventional methods. The polypeptide also exhibits high heat stability, thus sterilization in autoclaves can be performed conveniently. The current sequence is that of a cell attachment-related auxiliary peptide of the invention.

XX CC

SQ Sequence 24 AA;

Query Match 100.0%; Score 44; DB 9; Length 24;

Best Local Similarity 84.6%; Pred. No. 3;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAXAA 13  
||| ||||| |||

Db 4 AAAAAAAAXAA 16

RESULT 39

ABG95821

ID ABG95821 standard; peptide; 25 AA.

XX AC ABG95821;

XX DT 09-DEC-2002 (first entry)

XX DE Transmembrane domain peptide #1.

XX KW Transmembrane domain; human; cow; papillomavirus; HIV; psoriasis; composite receptor; cell proliferative disorder; cystic fibrosis; T-cell lymphotropic virus; human immunodeficiency virus; cytostatic; antipsoriatic; respiratory.

XX OS Unidentified.

XX PN WO200262823-A2.

XX PD 15-AUG-2002.

XX PF 01-FEB-2002; 2002WO-US002837.

XX PR 02-FEB-2001; 2001US-0265624P.

XX PA (UYVA) UNIV YALE.

XX PI Rabinovich P, Bray-Ward P, Ward DC;

XX DR WPI; 2002-723154/78.

XX CC New isolated peptide and fusion peptide, useful for inserting proteins or peptides and nucleic acids into the cell membrane in vivo or in vitro that may be useful in treating cancer or other proliferative disorders, e.g. psoriasis.

XX PS Claim 5; Page 62; 111pp; English.

XX CC The invention relates to an isolated peptide used in compositions for expressing a composite receptor on the cell surface. The peptide is a fusion peptide comprising at least one cell penetrating domain which is linked to at least one transmembrane domain or to an adapter domain capable of binding to a molecule, where the orientation of the cell penetrating domain is independent of the orientation of the transmembrane or adapter domain. The peptides are useful for inserting proteins or peptides such as regulatory factors, enzymes, antibodies, drugs or toxins and nucleic acids such as DNA or RNA into the cell membrane either in vivo or in vitro, that may be useful in inhibiting or treating diseases or disorders, which could benefit from the addition of surface receptors to cells, such as cancer and other proliferative disorders like psoriasis and cystic fibrosis. The methods are useful for identifying agents that modulate the activity of a receptor protein and for preparing the fusion

```

CC polymer (II) is useful for modifying and improving performance of IR-
CC sensitive devices, IR sensors, IR filters, night telescopes and
CC thermosensitive detectors. (II) can be used for preparing films,
CC membranes or coatings that absorb specific wavelengths of infrared
CC radiation, and in optical applications e.g., as matrices to align non-
CC linear chromophores which are useful in creating materials for second
CC order non-linear orbits. (II) is useful in hydrogen catalysis, as
CC coatings for biomaterials, scaffolds for tissue engineering,
CC ferroelectric materials, artificial muscles, switching devices, etc. This
CC sequence represents a peptide used in the method of the invention.
XX
SQ Sequence 26 AA;

Query Match 100.0%; Score 44; DB 7; Length 26;
Best Local Similarity 84.6%; Pred. NO. 3.2;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAXAAAAAAAAAXA 13
   || ||||| ||
Db 4 AAAAAAAAAAAAAA 16

Search completed: September 9, 2006, 22:48:11
Job time : 99.4177 secs

```



GenCore version 5.1.9

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:48:41 ; Search time 14.8101 Seconds  
(without alignments)  
84.457 Million cell updates/sec

Title: US-10-617-568-4

Perfect score: 54

Sequence: 1 AAFAAAAA 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR\_80:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	90.7	465	G02738	FRAC-4 - human
2	47	87.0	530	I38558	Mi-2 autoantigen 2
3	47	87.0	570	S72227	finger protein sob
4	47	87.0	892	T09071	SH3 domains-contai
5	47	87.0	1533	A46221	abdominal segment
6	47	87.0	2038	A43742	female sterile hom
7	46	85.2	150	T12547	hypothetical prote
8	46	85.2	220	JC5954	ribosomal protein
9	46	85.2	233	S11563	probable MASH-2 pr
10	46	85.2	238	A48279	achaete scute prot
11	46	85.2	266	S61522	LMO protein - frui
12	46	85.2	289	A43562	homeotic protein H
13	46	85.2	305	I57039	genomic screen hom
14	46	85.2	314	JC5273	paired type homeob
15	46	85.2	323	SL6318	homeotic protein H
16	46	85.2	331	B47236	zinc-finger protei
17	46	85.2	333	A39065	homeotic protein E
18	46	85.2	334	G02409	protein kinase C-b
19	46	85.2	364	I48188	gene Nkx6.1 protei
20	46	85.2	374	T03875	probable homeobox
21	46	85.2	375	T03874	probable homeobox
22	46	85.2	378	A44443	basic helix-loop-h
23	46	85.2	401	A48423	engrailed homeodom
24	46	85.2	403	A53662	homeotic protein H
25	46	85.2	420	T14911	bZIP DNA-binding p
26	46	85.2	443	I38239	transcription fact
27	46	85.2	475	A43915	homeotic protein e
28	46	85.2	477	A47236	zinc-finger protei
29	46	85.2	482	JC7583	basic helix-loop-h

30	46	85.2	490	2	T09084	phosphatidylinosit
31	46	85.2	494	2	A42170	zinc finger protei
32	46	85.2	495	1	S31223	transcription fact
33	46	85.2	497	2	JC5076	myc-associated zin
34	46	85.2	513	2	A48233	polyomavirus enhan
35	46	85.2	552	1	WJPEEN	homeotic protein e
36	46	85.2	627	2	T02610	probable YME1 ATP-
37	46	85.2	644	2	S39356	transcription fact
38	46	85.2	779	1	S40382	box A-binding fact
39	46	85.2	796	2	JC7555	Cl4orf4 protein -
40	46	85.2	805	2	T49385	hypothetical prote
41	46	85.2	806	2	T13690	hypothetical prote
42	46	85.2	828	2	C88402	protein H05C05.1 [
43	46	85.2	829	2	A34692	ecdysone-induced p
44	46	85.2	873	2	B53225	ecdysone-induced p
45	46	85.2	883	2	S40722	puff 74E protein -
46	46	85.2	1028	2	A56038	DNA-binding protei
47	46	85.2	1065	2	T13230	dachshund isoform
48	46	85.2	1072	2	T13232	dachshund protein
49	46	85.2	1074	2	T13229	dachshund protein
50	46	85.2	1081	2	T13231	dachshund protein
51	46	85.2	1180	2	S69205	stripe a/b protein
52	46	85.2	1205	2	A55015	bumetanide-sensiti
53	46	85.2	1212	2	A57187	bumetanide-sensiti
54	46	85.2	1213	2	S16356	ovo protein - frui
55	46	85.2	1355	2	S40022	spalt protein - fr
56	46	85.2	1506	2	T51900	related to PAN2 pr
57	46	85.2	1668	2	T13748	sex comb protein -
58	46	85.2	1891	2	T13594	hypothetical prote
59	46	85.2	2639	2	T31328	fibroin - Chinese
60	45	83.3	451	2	D88395	protein F53A3.6 [i
61	45	83.3	2175	1	S03170	homeotic protein c
62	44	81.5	377	2	T04213	heat shock transcr
63	44	81.5	470	2	S33639	finger protein esc
64	43	79.6	171	2	S78525	alpha-amylase inhi
65	43	79.6	296	2	S45336	finger protein, Sp
66	43	79.6	337	2	S06956	segmentation prote
67	43	79.6	361	1	OPBYC	cytochrome-c perox
68	43	79.6	392	2	B48423	homeotic protein e
69	43	79.6	405	2	T12117	26S proteasome reg
70	43	79.6	448	2	S17370	DNA-binding protei
71	43	79.6	604	2	A39369	homeotic protein B
72	43	79.6	606	2	S13367	Om (1b) protein - f
73	43	79.6	640	2	A41726	homeotic protein B
74	43	79.6	642	2	S27806	homeotic protein B
75	43	79.6	748	2	T49633	glucan 1,4-alpha-g
76	43	79.6	832	2	A31246	neurogenic protein
77	43	79.6	833	2	S19087	gene Delta protein
78	43	79.6	880	2	S00670	neurogenic repetit
79	43	79.6	1077	2	A44067	serine-rich protei
80	43	79.6	1166	2	H86341	hypothetical prote
81	43	79.6	2115	2	T13049	eyelid - fruit fly
82	43	79.6	2957	2	T31152	hypothetical prote
83	43	79.6	4845	2	T31067	BiR repeat contain
84	43	79.6	5327	2	T13564	microtubule-associ
85	42	77.8	40	2	S58853	homeotic protein u
86	42	77.8	80	2	T31267	4-oxalocrotonate
87	42	77.8	85	1	FDFL4W	antifreeze protein
88	42	77.8	91	2	A22592	antifreeze protein
89	42	77.8	97	2	S02376	antifreeze protein
90	42	77.8	172	2	S35568	sex-determining pr
91	42	77.8	219	2	I51382	achaete-scute homo
92	42	77.8	231	2	S28186	achaete-scute locu
93	42	77.8	306	1	S59863	polyA binding prot
94	42	77.8	311	2	T45683	hypothetical prote
95	42	77.8	317	1	R5MS10	acidic ribosomal p
96	42	77.8	328	2	AB1781	hypothetical prote
97	42	77.8	376	2	A26066	segmentation prote
98	42	77.8	425	1	JH0710	transcription fact
99	42	77.8	425	2	S38153	paired box protein
100	42	77.8	509	2	A36392	segmentation prote

## ALIGNMENTS

```
RESULT 1
G02738
FRAC-4 - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Oct-2004
C:Accession: G02738
R:Enserback, S.
Submitted to the EMBL Data Library, June 1996
A:Reference number: H01646
A:Accession: G02738
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-465 <ENE>
A:Cross-references: UNIPROT:Q16676; UNIPARC:UPI000012ADC8; EMBL:U59832; NID:g1399238; PI
C:Genetics:
A:Gene: FRAC-4
F:125-216/Domain: fork head DNA-binding domain homology <FHD>

Query Match      90.7%; Score 49; DB 2; Length 465;
Best Local Similarity 92.3%; Pred. No. 8.4;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
   |||
Db 290 ALFAAAAA 302

RESULT 2
I38558
Mi-2 autoantigen 240 kDa protein - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 05-Nov-1999
C:Accession: I38558
J:Ge, Q.; Nilasena, D.S.; O'Brien, C.A.; Frank, M.B.; Targoff, I.N.
J. Clin. Invest. 96, 1730-1737, 1995
A:Title: Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 au
A:Reference number: I38558; MUID:96013633; PMID:7560064
A:Accession: I38558
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-530 <RES>
A:Cross-references: UNIPARC:UPI000016A09A; EMBL:U08379; NID:g761717; PIDN:AAC50228.1; PI

Query Match      87.0%; Score 47; DB 2; Length 530;
Best Local Similarity 92.3%; Pred. No. 17;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
   |||
Db 87 AAVAAAAA 99

RESULT 3
S72227
finger protein sob - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 04-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: S72227
R:Hart, M.C.; Wang, L.; Coulter, D.E.
Genetics 144, 171-182, 1996
A:Title: Comparison of the structure and expression of odd-skipped and two related genes
A:Reference number: S72227; MUID:97032935; PMID:8878683
A:Accession: S72227
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-577 <HAR>
A:Cross-references: UNIPROT:Q24571; UNIPARC:UPI000007A28D; EMBL:U62004; NID:g1480193; PI
C:Genetics:
A:Gene: sob
A:Cross-references: FlyBase:FBgn0004892
```

C;Keywords: zinc finger

Query Match 87.0%; Score 47; DB 2; Length 577;  
Best Local Similarity 92.3%; Pred. No. 18;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13  
 |||  
Db 273 AAVAAAAA 285

## RESULT 4

T09071  
SH3 domains-containing protein POSH - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T09071  
R:Tapon, N.; Nagata, K.; Lamarche, N.; Hall, A.  
EMBO J. 17, 1395-1404, 1998  
A:Reference number: Z16552; MUID:98151363; PMID:9482736  
A:Accession: T09071  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-892 <TAP>  
A:Cross-references: UNIPROT:O70254; UNIPARC:UPI0000027BF6; EMBL:AF030131; NID:g3002587;  
A:Experimental source: cell line Ras-transformed NIH 3T3 cells  
A:Note: activates JNK/SAPK cascade; Rac-binding protein  
C:Genetics:  
A:Gene: POSH  
C:Keywords: signal transduction  
F:8-58/Domain: RING finger homology <RRN>

Query Match 87.0%; Score 47; DB 2; Length 892;  
Best Local Similarity 92.3%; Pred. No. 26;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13  
 |||  
Db 419 AAVAAAAA 431

## RESULT 5

A46221  
abdominal segment formation protein pumilio - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: A46221; S22026  
R:Barker, D.D.; Wang, C.; Moore, J.; Dickinson, L.K.; Lehmann, R.  
Genes Dev. 6, 2312-2326, 1992  
A:Title: Pumilio is essential for function but not for distribution of the Drosophila a  
A:Reference number: A46221; MUID:93093466; PMID:1459455  
A:Contents: embryo  
A:Accession: A46221  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1533 <BAR>  
A:Cross-references: UNIPROT:P25822; UNIPARC:UPI000016BD14; GB:L07943; NID:g158190; PID:g  
R:Macdonald, P.M.  
submitted to the EMBL Data Library, October 1991  
A:Reference number: S22026  
A:Accession: S22026  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-361, 'A', 363-1102, 'R', 1104-1405, 'KN', 1408-1495, 'V', 1497-1518, 'S', 1520-1533  
A:Cross-references: UNIPARC:UPI00001329FD; EMBL:X62589; NID:g83393; PID:g8394  
C:Genetics:  
A:Gene: FlyBase:pum  
A:Cross-references: FlyBase:FBgn0003165

Query Match 87.0%; Score 47; DB 2; Length 1533;  
Best Local Similarity 92.3%; Pred. No. 40;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFRAAAAAAAAAA 13  
|||  
Db 1050 AAVAAAAAAAAA 1062

## RESULT 6

A43742  
female sterile homeotic protein, 205K - fruit fly (Drosophila melanogaster)  
N:Alternate names: membrane protein fish, 205K  
N:Contains: female sterile homeotic protein, 110K  
C:Species: Drosophila melanogaster  
C:Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 31-Dec-2004  
C:Accession: A43742; B43742  
R:Haynes, S.R.; Mozer, B.A.; Bhatia-Dey, N.; Dawid, I.B.  
Dev. Biol. 134, 246-257, 1989  
A:Title: The Drosophila fish locus, a maternal effect homeotic gene, encodes apparent mem  
A:Reference number: A43742; MUID:89276730; PMID:2567251  
A:Accession: A43742  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2038 <HAY>  
A:Cross-references: UNIPROT:P13709; UNIPARC:UPI000012AC6C; EMBL:M23221; NID:g157452; PID  
A:Accession: B43742  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1106 <HA2>  
A:Cross-references: UNIPARC:UPI000017A10F; EMBL:M23222  
C:Genetics:  
A:Gene: fish  
A:Cross-references: FlyBase:FBgn0004656  
C:Keywords: alternative splicing; transmembrane protein  
F:1-2038/Product: female sterile homeotic protein, 205K #status predicted <MA2>  
F:1-1106/Product: female sterile homeotic protein, 110K #status predicted <MAT>  
F:59-116/Domain: bromodomain homology <BRO1>  
F:503-560/Domain: bromodomain homology <BRO2>

Query Match 87.0%; Score 47; DB 2; Length 2038;  
Best Local Similarity 92.3%; Pred. No. 50;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFRAAAAAAAAAA 13  
|||  
Db 330 AAVAAAAAAAAA 342

## RESULT 7

T12547  
hypothetical protein DKFZp586E1621.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T12547  
R:Ottenwälder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: T12547  
A:Accession: T12547  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-150 <ORT>  
A:Cross-references: UNIPROT:Q9Y4M1; UNIPARC:UPI0000071DDD; EMBL:AL080235  
A:Experimental source: adult uterus; clone DKFZp586E1621  
C:Genetics:  
A>Note: DKFZp586E1621.1

Query Match 85.2%; Score 46; DB 2; Length 150;  
Best Local Similarity 92.3%; Pred. No. 8.4;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFRAAAAAAAAAA 13  
|||  
Db 129 AAVAAAAAAAAA 141

## RESULT 8

JC5954  
ribosomal protein L14 - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 02-Sep-2000  
C:Accession: JC5954  
R:Tanaka, M.; Tanaka, T.; Harata, M.; Suzuki, T.; Mitsui, Y.  
Biochem. Biophys. Res. Commun. 243, 531-537, 1998  
A:Title: Triplet repeat-containing ribosomal protein L14 gene in immortalized human endo  
A:Reference number: JC5954; MUID:98153799; PMID:9480843  
A:Accession: JC5954  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-220 <TAN>  
A:Cross-references: UNIPARC:UPI000004EG22; DDBJ:D87735; NID:g1620021; PIDN:BAAL13443.1; P  
C:Superfamily: rat ribosomal protein L14

Query Match 85.2%; Score 46; DB 2; Length 220;  
Best Local Similarity 92.3%; Pred. No. 11;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFRAAAAAAAAAA 13  
|||  
Db 150 AAVAAAAAAAAA 162

## RESULT 9

S11563  
probable MASH-2 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S11563  
R:Johnson, J.E.; Birren, S.J.; Anderson, D.J.  
Nature 346, 858-861, 1990  
A:Title: Two rat homologues of Drosophila achaete-scute specifically expressed in neuro  
A:Reference number: S11562; MUID:90363294; PMID:2392153  
A:Accession: S11563  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-233 <JOH>  
A:Cross-references: UNIPROT:P19359; UNIPARC:UPI0000047567; EMBL:X53725; NID:g56630; PID  
C:Keywords: DNA binding

Query Match 85.2%; Score 46; DB 2; Length 233;  
Best Local Similarity 92.3%; Pred. No. 12;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFRAAAAAAAAAA 13  
|||  
Db 32 AAVAAAAAAAAA 44

## RESULT 10

A48279  
achaete scute protein - human  
C:Species: Homo sapiens (man)  
C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 09-Jul-2004  
C:Accession: A48279  
R:Ball, D.W.; Azzoli, C.G.; Baylin, S.B.; Chi, D.; Dou, S.; Donis-Keller, H.; Kumarasw  
Proc. Natl. Acad. Sci. U.S.A. 90, 5648-5652, 1993  
A:Title: Identification of a human achaete-scute homolog highly expressed in neuroendoc  
A:Reference number: A48279; MUID:93296195; PMID:8390674  
A:Accession: A48279  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-238 <RES>  
A:Cross-references: UNIPROT:P50553; UNIPARC:UPI0000047566; GB:L08424; NID:g306459; PID:g

Query Match 85.2%; Score 46; DB 2; Length 238;  
Best Local Similarity 92.3%; Pred. No. 12;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFRAAAAAAAAAA 13

```
Db      35 AAAAAAAAAAAAAA 47
|||||
A;Residues: 1-305 <RES>
A;Cross-references: UNIPARC:UPI0000024E69; GB:S79041; NID:g1042008; PIDN:AAB334947.1; PID
C;Genetics:
A;Gene: Gsh-2
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:204-260/Domain: homeobox homology <HOX>

Query Match      85.2%; Score 46; DB 2; Length 305;
Best Local Similarity 92.3%; Pred. No. 15;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAFAAAAAAAAAAAA 13
      |||
Db      147 AAAAAAAAAAAAAA 159

RESULT 14
JC5273
paired type homeobox protein, NBP - human
C;Species: Homo sapiens (man)
C;Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 05-Oct-2004
C;Accession: JC5273
R;Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K.
DNA Res. 3, 311-320, 1996
A;Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific
A;Reference number: JC5272; MUID:97191543; PMID:9039501
A;Contents: neuroblastoma cell
A;Accession: JC5273
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-314 <YOK>
A;Cross-references: UNIPROT:Q99453; UNIPARC:UPI0000131D16; DBJ:D82344; NID:g1841337; PID
C;Comment: This protein is a transcriptional repressor involved in regulating gene expression
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:99-155/Domain: homeobox homology <HOX>

Query Match      85.2%; Score 46; DB 2; Length 314;
Best Local Similarity 92.3%; Pred. No. 15;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AAFAAAAAAAAAAAA 13
      |||
Db      241 AAAAAAAAAAAAAA 253

RESULT 15
S16318
homeotic protein Hox 4.6 - mouse
N;Alternate names: homeotic protein Hoxd-11
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Oct-2004
C;Accession: S16318; S40403; S57443
R;Izpisua-Belmonte, J.C.; Falkenstein, H.; Dolle, P.; Renucci, A.; Duboule, D.
EMBO J. 10, 2279-2289, 1991
A;Title: Murine genes related to the Drosophila AbdB homeotic gene are sequentially expressed
A;Reference number: S16317; MUID:91293104; PMID:1676674
A;Accession: S16318
A;Molecule type: DNA
A;Residues: 1-323 <IZP>
A;Cross-references: UNIPROT:P23813; UNIPARC:UPI0000029F10; EMBL:X60761; NID:g51420; PIDN:
R;Gerard, M.; Duboule, D.; Zakany, J.
EMBO J. 12, 3539-3550, 1993
A;Title: Structure and activity of regulatory elements involved in the activation of the
A;Reference number: S40403; MUID:94074553; PMID:7902810
A;Accession: S40403
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-323 <GER>
A;Cross-references: UNIPARC:UPI0000029F10; EMBL:X71422; NID:g397508; PIDN:CAAS0553.1; PID
R;Takada, S.; Cook, M.; Kramlauf, R.; McMahon, A.P.
submitted to the EMBL Data Library, May 1991
A;Description: Genomic sequence of mouse Hox-4.6.
A;Reference number: S57443
```

```
A:Accession: S57443
A:Molecule type: DNA
A:Residues: 'MADFECGSAAS', 1-323 <TAK>
A:Cross-references: UNIPARC:UPI0000026A38; EMBL:X60395; NID:g871427; PIDN:CAA42943.1; PI
C:Genetics:
A:Gene: Hoxd-11
A:Introns: 246/1
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:252-308/Domain: homeobox homology <HOX>

Query Match      85.2%; Score 46; DB 2; Length 323;
Best Local Similarity 92.3%; Pred. No. 15;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
Db 93 AAAAAAAAAAAAAA 105

RESULT 16
B47236
A:Accession: B47236
A:Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates
A:Reference number: A47236; MUID:93087555; PMID:1454839
A:Accession: B47236
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-331 <KEN>
A:Cross-references: UNIPARC:UPI000017C615
A:Experimental source: insulinoma cell line T
A>Note: sequence extracted from NCBI backbone (NCBIP:119832)

Query Match      85.2%; Score 46; DB 2; Length 331;
Best Local Similarity 92.3%; Pred. No. 16;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
Db 83 AAAAAAAAAAAAAA 95

RESULT 17
A39065
A:Accession: A39065
A:Title: EVX2, a human homeobox gene homologous to the even-skipped segmentation gene,
A:Reference number: A39065; MUID:91257849; PMID:1675198
A:Accession: A39065
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-333 <DAE>
A:Cross-references: UNIPARC:UPI000017A29D
A:Gene: GDB:EVX2
A:Cross-references: GDB:127528; OMIM:142991
A:Map position: 2q24.3-2q31
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:46-102/Domain: homeobox homology <HOX>

Query Match      85.2%; Score 46; DB 2; Length 333;
Best Local Similarity 92.3%; Pred. No. 16;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
```

```
Db 213 AASAAAAAAAAA 225

RESULT 18
G02409
A:Accession: G02409
A:Title: protein kinase C-binding protein RACK17 - human (fragment)
A:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C:Accession: G02409
R:Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Kikkawa, U.
submitted to the EMBL Data Library, February 1996
A:Reference number: H01212
A:Accession: G02409
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-334 <KUR>
A:Cross-references: UNIPARC:UPI000016A2D0; EMBL:U48250; NID:g1199656; PIDN:AAC72247.1;

Query Match      85.2%; Score 46; DB 2; Length 334;
Best Local Similarity 92.3%; Pred. No. 16;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
Db 236 AAAAAAAAAAAAAA 248

RESULT 19
I48188
A:Accession: I48188
A:Title: gene Nkx6.1 protein - golden hamster
A:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Oct-2004
C:Accession: I48188
R:Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994
A:Title: Pancreatic beta cells express a diverse set of homeobox genes.
A:Reference number: I48185; MUID:95083670; PMID:7991607
A:Accession: I48188
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-364 <RES>
A:Cross-references: UNIPROT:Q60554; UNIPARC:UPI00001301FA; EMBL:X81409; NID:g587466; PI
C:Genetics:
A:Gene: NKx6.1
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:237-293/Domain: homeobox homology <HOX>

Query Match      85.2%; Score 46; DB 2; Length 364;
Best Local Similarity 92.3%; Pred. No. 17;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
Db 137 AAAAAAAAAAAAAA 149

RESULT 20
T03875
A:Accession: T03875
A:Title: probable homeobox protein OSH45, splice form OSH45 [similarity] - rice
A:Species: Oryza sativa (rice)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-Oct-2004
C:Accession: T03875
R:Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Ma
Plant J. 7, 927-938, 1995
A:Title: Alternative RNA products from a rice homeobox gene.
A:Reference number: Z15126; MUID:95322999; PMID:7599652
A:Accession: T03875
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-374 <TAM>
A:Cross-references: UNIPROT:P93423; UNIPARC:UPI00000AB610; EMBL:D49704; NID:g1805615; PI
A:Experimental source: cv. Nipponbare
```

C;Comment: For alternative splice forms, see PIR:T03874.

C;Genetics:

A;Gene: H45

A;Introns: 174/3; 215/1; 277/3; 321/3; 367/2

C;Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus

F;291-352/Domain: homeobox homology <HOX>

Query Match85.2%; Score 46; DB 2; Length 374;Best Local Similarity92.3%; Pred. No. 17;Matches12; Conservative0; Mismatches1; Indels0; Gaps0;

Qy1AAFAAAAAAAAAA13|||||

Db64AAAAAAAAAAAA76

RESULT 21

T03874

probable homeobox protein OSH45, splice form OSH44 [similarity] - rice

N;Contains: probable homeobox protein OSH45, splice form OSH42

C;Species: Oryza sativa (rice)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Oct-2004

C;Accession: T03874; T03876

R;Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Mat Plant J. 7, 927-938, 1995

A;Title: Alternative RNA products from a rice homeobox gene.

A;Reference number: 215126; MUID:95322999; PMID:7599652

A;Accession: T03874

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-375 <TAM1>

A;Cross-references: UNIPROT:P93424; UNIPARC:UPI00000AASD5; EMBL:D49704; NID:g1805615; PI

A;Experimental source: cv. Nipponbare; splice form OSH44

A;Accession: T03876

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 197-367 <TAM2>

A;Cross-references: UNIPARC:UPI000016DEBB; EMBL:D49704; NID:g1805615; PIDN:BAA08554.1; F

A;Experimental source: cv. Nipponbare; splice form OSH42

C;Comment: For an alternative splice form, see PIR:T03875.

C;Genetics:

A;Gene: H45

A;Introns: 174/3; 215/1; 277/3; 321/3; 367/2

C;Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus

F;1-375/Product: probable homeobox protein OSH45, splice form OSH44 #status predicted <M

F;197-367/Product: probable homeobox protein OSH45, splice form OSH42 #status predicted <H

F;291-352/Domain: homeobox homology <HOX>

Query Match85.2%; Score 46; DB 2; Length 375;Best Local Similarity92.3%; Pred. No. 17;Matches12; Conservative0; Mismatches1; Indels0; Gaps0;

Qy1AAFAAAAAAAAAA13|||||

Db64AAAAAAAAAAAA76

RESULT 22

A4443

basic helix-loop-helix transcription repressor H - fruit fly (Drosophila virilis)

C;Species: Drosophila virilis

C;Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Jul-1998

C;Accession: A4443

R;Wainwright, S.M.; Ish-Horowicz, D. Mol. Cell. Biol. 12, 2475-2483, 1992

A;Title: Point mutations in the Drosophila hairy gene demonstrate in vivo requirements f

A;Reference number: A4443; MUID:92269819; PMID:1588951

A;Accession: A4443

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-378 <WAI>

A;Cross-references: UNIPARC:UPI000017BEF1

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI backbone (NCBIN:102964, NCBIIP:102966)

C;Genetics:

A;Gene: FlyBase:Dvir/h

A;Cross-references: FlyBase:FBgn0013115

Query Match85.2%; Score 46; DB 2; Length 378;Best Local Similarity92.3%; Pred. No. 18;Matches12; Conservative0; Mismatches1; Indels0; Gaps0;

Qy1AAFAAAAAAAAAA13|||||

Db269AAAAAAAAAAAA281

RESULT 23

A48423

engrailed homeodomain-containing protein En-1 - mouse

N;Alternate names: homeotic protein En-1

C;Species: Mus musculus (house mouse)

C;Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004

C;Accession: A48423; SI3009; A26629; A24778

R;Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.L. Dev. Genet. 13, 345-358, 1992

A;Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene

A;Reference number: A48423; MUID:93185339; PMID:1363401

A;Accession: A48423

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-401 <LOG>

A;Cross-references: UNIPROT:P09065; UNIPARC:UPI00000299A1

A;Experimental source: CD-1, embryo

A;Note: sequence extracted from NCBI backbone (NCBIIP:126620)

R;Holland, P.W.H.; Williams, N.A. PNAS Lett. 277, 250-252, 1990

A;Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution.

A;Reference number: SI3009; MUID:91099509; PMID:1980115

A;Accession: SI3009

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 321-380 <HOL>

A;Cross-references: UNIPARC:UPI000017A2AF

R;Joyner, A.L.; Martin, G.R. Genes Dev. 1, 29-38, 1987

A;Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrai

A;Reference number: A91620; MUID:88112776; PMID:2892757

A;Accession: A26629

A;Molecule type: DNA; mRNA

A;Residues: 278-401 <JOY>

A;Cross-references: UNIPARC:UPI000016CA7C; GB:Y00201; GB:M11987; NID:g49587; PIDN:CAA68

R;Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R. Cell 43, 29-37, 1985

A;Title: Expression during embryogenesis of a mouse gene with sequence homology to the

A;Reference number: A24778; MUID:86079501; PMID:2416459

A;Accession: A24778

A;Molecule type: DNA

A;Residues: 311-401 <JO2>

A;Cross-references: UNIPARC:UPI000017A2E0

C;Genetics:

A;Gene: en.1

A;Map position: 1

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation

F;313-369/Domain: homeobox homology <HOX>

Query Match85.2%; Score 46; DB 2; Length 401;Best Local Similarity92.3%; Pred. No. 18;Matches12; Conservative0; Mismatches1; Indels0; Gaps0;

Qy1AAFAAAAAAAAAA13|||||

Db207AAAAAAAAAAAA219

RESULT 24

A53662  
homeotic protein HB9 - human  
C/Species: Homo sapiens (man)  
C/Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 31-Dec-2004  
C/Accession: A53662  
R/Harrison, K.A.; Druey, K.M.; Deguchi, Y.; Tusciano, J.M.; Kehrl, J.H.  
J. Biol. Chem. 269, 19968-19975, 1994  
A/Title: A novel human homeobox gene distantly related to proboscipedia is expressed in  
A/Reference number: A53662; MUID:94327547; PMID:7914194  
A/Accession: A53662  
A/Status: preliminary; not compared with conceptual translation  
A/Molecule type: DNA  
A/Residues: 1-403 <HAR>  
A/Cross-references: UNIPARC:UPI000017A29E; GB:U07663  
A/Note: the nucleotide sequence and conceptual translation as given are self-consistent  
C/Genetics:  
A/Genes: GDB:HLXB9  
A/Cross-references: GDB:I36411; OMIM:142994  
A/Map position: 1q41-1q42.1  
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F/244-300/Domain: homeobox homology <HOX>

Query Match 85.2%; Score 46; DB 2; Length 403;  
Best Local Similarity 92.3%; Pred. No.18;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
|| |||||  
Db 121 AAAAAAAAAAAAAA 133

RESULT 25  
T14911  
bZIP DNA-binding protein - parsley  
C/Species: Petroselinum crispum (parsley)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 31-Dec-2004  
C/Accession: T14911  
R/Kircher, S.; Ledger, S.; Hayashi, H.; Weisshaar, B.; Schafer, E.; Frohnmeyer, H.  
Mol. Gen. Genet. 257, 595-605, 1998  
A/Title: CPRF4a, a novel plant bZIP protein of the CPRF family: comparative analysis of  
A/Reference number: 218261; MUID:98265918; PMID:9604882  
A/Accession: T14911  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-420 <KIR>  
A/Cross-references: UNIPROT:O82037; UNIPARC:UPI00000A12C8; EMBL:Y10810; NID:G3336904; P:  
C/Genetics:  
A/Note: CPRF4b  
C/Superfamily: bZIP G-box-binding protein; fos/jun DNA-binding domain homology  
C/Keywords: DNA binding; leucine zipper

Query Match 85.2%; Score 46; DB 2; Length 420;  
Best Local Similarity 92.3%; Pred. No.19;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
|| |||||  
Db 31 AAAAAAAAAAAAAA 43

RESULT 26  
I38239  
transcription factor SOX3 - human  
N/Alternate names: SRY (sex determining region Y)-box 3  
C/Species: Homo sapiens (man)  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: I38239; I38242; S67816  
R/Stevanovic, M.; Lovell-Badge, R.; Collignon, J.; Goodfellow, P.N.  
Hum. Mol. Genet. 2, 2013-2018, 1993  
A/Title: SOX3 is an X-linked gene related to SRY.  
A/Reference number: I38239; MUID:94154672; PMID:8111369  
A/Accession: I38239  
A/Status: preliminary; translated from GB/EMBL/DBJ

```

Query Match      85.2%; Score 46; DB 2; Length 477;
Best Local Similarity 92.3%; Pred. No. 21;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFASAAAAA 13
   ||| ||| ||| ||| |||
DB 96 AAAAAA 108

RESULT 29
JC7583
basic helix-loop-helix protein, DEC2 - human
C;Species: Homo sapiens (man)
C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 09-Jul-2004
C;Accession: JC7583
R;Fujimoto, K.; Shen, M.; Noshiro, M.; Matsubara, K.; Shingu, S.; Honda, K.; Yoshida, E.
Biochem. Biophys. Res. Commun. 280, 164-171, 2001
A;Title: Molecular cloning and characterization of DEC2, a new member of basic helix-loop-
A;Reference number: JC7583; MUID:21092582; PMID:11162494
A;Accession: JC7583
A;Molecule type: mRNA
A;Residues: 1-482 <FULL>
A;Cross-references: UNIPROT:Q9C0J9; UNIPARC:UPI0000052B52; DDBJ:AB044088
C;Comment: This protein, a novel member of the DEC subfamily of basic helix-loop-helix P
C;Genetics:
A;Gene: dec2
A;Map position: 12p11.23-p12.1
A;Keywords: transcription factor
F;1-173/Region: highly conserved #status predicted
F;130-173/Domain: Orange #status predicted <ORA>
F;286-411/Region: alanine and glycine-rich #status predicted

Query Match      85.2%; Score 46; DB 2; Length 482;
Best Local Similarity 92.3%; Pred. No. 21;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFASAAAAA 13
   ||| ||| ||| ||| |||
DB 396 AAAAAA 408

RESULT 30
T09084
phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)
C;Species: Chlamydomonas reinhardtii
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: T09084
R;Molendijk, A.J.; Irvine, R.F.
Plant Mol. Biol. 37, 53-66, 1998
A;Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinosit
A;Reference number: Z16411; MUID:98281574; PMID:9620264
A;Accession: T09084
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-490 <MOL>
A;Cross-references: UNIPROT:O04270; UNIPARC:UPI000009C6FA; EMBL:U97663; NID:G2109290; PT
A;Experimental source: strain cw-15
C;Genetics:
A;Introns: 265/3; 331/3; 370/3; 455/1; 481/3

Query Match      85.2%; Score 46; DB 2; Length 490;
Best Local Similarity 92.3%; Pred. No. 22;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFASAAAAA 13
   ||| ||| ||| ||| |||
DB 115 AAAAAA 127

RESULT 31
A42170
zinc finger protein MAZ - human (fragment)
N;Alternate names: MYC-associated zinc finger protein MAZ; zinc finger protein ZF87

```



```
QY      1  AAFAAAAAAAAA 13
Db      186 AAAAAAAAAA 198

RESULT 33
JCS076
myc-associated zinc-finger protein - human
N:Alternate names: MAZ protein
C:Species: Homo sapiens (man)
C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 05-Nov-1999
A:Accession: JCS076
A:Residues: 1-497 <TSU>
R:Tsutsui, H.; Sakatsune, O.; Itakura, K.; Yokoyama, K.K.
Biochem. Biophys. Res. Commun. 226, 801-809, 1996
A:Title: Members of the MAZ family: A novel cDNA clone for MAZ from human pancreatic islet
A:Reference number: JCS076; MUID:96428591; PMID:9831693
A:Accession: JCS076
A:Molecule type: mRNA
A:Residues: 1-497 <TSU>
A:Cross-references: UNIPARC:UPI0000163B39; DDBJ:D85131; NID:gl752741; PIDN:BAAL2728.1; E
A:Experimental source: pancreatic islet
C:Comment: This protein plays a role in the control of transcriptional initiation of gen
and between the introns of the mouse gene for immunoglobulin M-D.
C:Keywords: phosphoprotein; zinc finger
F:145,204,480/Binding site: phosphate (Ser) (covalent) #status predicted
F:345/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match      85.2%; Score 46; DB 2; Length 497;
Best Local Similarity 92.3%; Pred. No. 22;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  AAFAAAAAAAAA 13
Db      122 AAAAAAAAAA 134

RESULT 34
A48233
polyomavirus enhancer-binding protein 2 alpha chain type 1 - mouse
N:Alternate names: PE22 alpha chain type 1; PE22 alpha chain type 2; PEBP2 alpha chain t
C:Species: Mus musculus (house mouse)
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Jul-2004
C:Accession: A48233; B48233
R:Ogawa, E.; Maruyama, M.; Kagoshima, H.; Inuzuka, M.; Lu, J.; Satake, M.; Shigesada, K.
Proc. Natl. Acad. Sci. U.S.A. 90, 6859-6863, 1993
A:Title: PEBP2/PE22 represents a family of transcription factors homologous to the produ
A:Reference number: A48233; MUID:93342088; PMID:8341710
A:Accession: A48233
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-513 <OGA>
A:Cross-references: UNIPROT:Q08775; UNIPARC:UPI000002B2F8; GB:D14636; NID:g391766; PIDN:
A:Accession: B48233
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-304,'L',306 <OG2>
A:Cross-references: UNIPARC:UPI000002B2F9; GB:D14637; NID:g391768; PIDN:BAAO3486.1; PID:
C:Genetics:
A:Gene: PEBP2alphaA
C:Superfamily: transcription factor CBF alpha 2
C:Keywords: alternative splicing; DNA binding; T-cell; transcription factor; transcripti

Query Match      85.2%; Score 46; DB 2; Length 513;
Best Local Similarity 92.3%; Pred. No. 22;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  AAFAAAAAAAAA 13
Db      64 AAAAAAAAAA 76

RESULT 35
```

```
WJFFEN
homeotic protein engrailed - fruit fly (Drosophila melanogaster)
N:Alternate names: specific body pattern development protein
C:Species: Drosophila melanogaster
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 09-Jul-2004
A:Accession: A90862; A93354; A03321; A03322; A25682; S03667
R:Poole, S.J.; Kauvar, L.M.; Drees, B.; Kornberg, T.
Cell 40, 37-43, 1985
A:Title: The engrailed locus of Drosophila: structural analysis of an embryonic transcr
A:Reference number: A90862; MUID:85099327; PMID:3917855
A:Accession: A90862
A:Molecule type: mRNA
A:Residues: 1-552 <POO>
A:Cross-references: UNIPROT:P02836; UNIPARC:UPI000012CA13; GB:M10017; NID:gl57363; PIDN:
R:Effosse, A.; McGinnis, W.J.; Gehring, W.J.
Nature 313, 284-289, 1985
A:Title: Isolation of a homeo box-containing gene from the engrailed region of Drosoph
A:Reference number: A93354; MUID:90114393; PMID:2481829
A:Accession: A93354
A:Molecule type: DNA
A:Residues: 447-485,'E',487-518,'WH' <FUO>
A:Cross-references: UNIPARC:UPI00001745BF; GB:X01765; GB:K03059; NID:g8084; PIDN:CAA259
A:Note: the translation of the nucleotide sequence from Fig. 5 is shown, not the transl
n in Fig. 6
R:Kassis, J.A.; Poole, S.J.; Wright, D.K.; O'Farrell, P.H.
EMBO J. 5, 3583-3589, 1986
A:Title: Sequence conservation in the protein coding and intron regions of the engrail
A:Reference number: A91059; MUID:87161768; PMID:2881781
A:Contents: annotation; intron locations and sequences
R:Gay, N.J.; Poole, S.J.; Kornberg, T.B.
Nucleic Acids Res. 16, 6637-6647, 1988
A:Title: The Drosophila engrailed protein is phosphorylated by a serine-specific protei
A:Reference number: S03667; MUID:88289425; PMID:2899884
A:Contents: annotation; potential phosphorylation sites; homeobox domain
C:Comment: This protein specifies the body segmentation pattern.
C:Genetics:
A:Gene: en
A:Cross-references: FlyBase:FBgn0000577
A:Map position: 2R,62.0 (48A1-4)
A:Introns: 438/1; 470/3
C:Superfamily: engrailed homeotic protein; homeobox homology
C:Keywords: DNA binding; embryo; homeobox; nucleus; segmentation; transcription regulat
F:26-53/Region: glutamine-rich
F:68-87/Region: alanine-rich
F:232-240/Region: alanine-rich
F:320-368/Region: serine-rich
F:455-511/Domain: homeobox homology <HOX>

Query Match      85.2%; Score 46; DB 1; Length 552;
Best Local Similarity 92.3%; Pred. No. 24;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  AAFAAAAAAAAA 13
Db      68 AAAAAAAAAA 80

RESULT 36
T02610
probable YME1 ATP-dependent proteinase [imported] - Arabidopsis thaliana
N:Alternate names: YTA11 protein homolog T19L18.5
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 31-Dec-2004
A:Accession: T02610; H84656
R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC T19L18 genomic sequence.
A:Reference number: Z14681
A:Accession: T02610
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-627 <ROU>
A:Cross-references: UNIPROT:O80983; UNIPARC:UPI000017885F; EMBL:AC004747; NID:g3413696;
```

A;Experimental source: cultivar Columbia  
M.;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: H84656  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-627 <STO>  
A;Cross-references: UNIPARC:UPI000017885F; GB:AE002093; NID:g3413700; PIDN:AAC31223.1; C;  
Genetics:  
A;Gene: T19L18.5; At2g26140  
A;Map position: 2  
A;Introns: 40/1; 119/1; 160/3; 319/3  
C;Superfamily: FtsH/SEC18/CDC48-type ATP-binding domain homology

Query Match	85.2%	Score 46;	DB 2;	Length 627;
Best Local Similarity	92.3%	Pred. No. 26;		
Matches 12;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy 1 AAFAAAAAAAAA 13  
Db 598 AAAAAAAAAA 610

RESULT 37

S39356  
transcription factor btd - fruit fly (*Drosophila* sp.)  
C;Species: *Drosophila* sp.  
C;Date: 18-Feb-1994 #sequence\_revision 01-Dec-1995 #text\_change 07-May-1999  
C;Accession: S39356  
R;Wimmer, E.A.; Jaecle, H.; Pfeifle, C.; Cohen, S.M.  
Nature 366, 690-694, 1993  
A;Title: A *Drosophila* homologue of human Spl is a head-specific segmentation gene.  
A;Reference number: S39356; MUID:94081952; PMID:8259212  
A;Accession: S39356  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-644 <WIM>  
A;Cross-references: UNIPARC:UPI0000124C17; EMBL:Z29361; NID:g441283; PID:g441284  
C;Genetics:  
A;Gene: FlyBase:btd  
A;Cross-references: FlyBase:FBgn0000233  
A;Introns: 245/2

Query Match	85.2%	Score 46;	DB 2;	Length 644;
Best Local Similarity	92.3%	Pred. No. 27;		
Matches 12;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy 1 AAFAAAAAAAAA 13  
Db 205 AAAAAAAAAA 217

RESULT 38

S40382  
box A-binding factor - fruit fly (*Drosophila melanogaster*)  
N;Alternate names: ABF; transcription factor dGATAB  
C;Species: *Drosophila melanogaster*  
C;Date: 13-Jan-1995 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: S40382  
R;Abel, T.; Michelson, A.M.; Maniatis, T.  
Development 119, 623-633, 1993  
A;Title: A *Drosophila* GATA family member that binds to Adh regulatory sequences is expressed in the developing eye.  
A;Reference number: S40382; MUID:94244465; PMID:8187633  
A;Accession: S40382  
A;Molecule type: mRNA  
A;Residues: 1-779 <ABE>  
A;Cross-references: UNIPROT:P52172; UNIPARC:UPI00001355F; EMBL:X76217; NID:g441491; PID:  
C;Comment: This transcriptional activator is the earliest known marker of the developing  
C;Genetics:

Search completed: September 9, 2006, 23:01:50  
Job time : 15.8101 secs

---

GenCore version 5.1.1.9

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:38:21 ; Search time 119.139 Seconds  
(without alignments)  
100.934 Million cell updates/sec

Title: US-10-617-568-4

Perfect score: 54

Sequence: 1 AAFAAAAAAAAA 13

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : UniProt\_7.2.1\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query Match	Length	DB ID	Description
1	50	92.6	163	2	Q7YTA3_9MYRI	Q7yca3 glomeris ma
2	50	92.6	269	2	Q9ARR7_ORYSA	Q9arr7 oryza sativ
3	49	90.7	305	2	Q8IMU7_DROME	Q8imu7 drosophila
4	49	90.7	314	1	RLA0_EIMTE	Q867y7 eimeria ten
5	49	90.7	399	2	Q5B2U7_EMENI	Q5b2u7 aspergillus
6	49	90.7	422	2	Q9W4E5_DROME	Q9w4e5 drosophila
7	49	90.7	456	1	FOXDI1_MOUSE	Q61345 mus musculu
8	49	90.7	456	2	Q3UOW8_MOUSE	Q3uow8 m 12 days e
9	49	90.7	465	1	FOXDI1_HUMAN	Q3us18 m 16 days e
10	48	88.9	167	2	Q6Z8T9_ORYSA	Q16676 homo sapien
11	48	88.9	361	2	Q4PEI8_USTMA	Q628t9 oryza sativ
12	48	88.9	484	1	ZETB8_MOUSE	Q4pei8 ustilago ma
13	48	88.9	484	2	Q3US18_MOUSE	Q8gi10 mus musculu
14	48	88.9	640	2	Q9VTW5_DROME	Q3us18 m 16 days e
15	47	87.0	339	2	Q9FYS1_HORVD	Q9vtw5 drosophila
16	47	87.0	378	2	Q9NXD6_HUMAN	Q9fysl hordeum vul
17	47	87.0	395	2	Q8OUS0_MOUSE	Q9nxd6 homo sapien
18	47	87.0	395	2	Q8BKCI_MOUSE	Q8ous0 mus musculu
19	47	87.0	395	2	Q3TSG4_MOUSE	Q8bkci m 0 day nec
20	47	87.0	395	2	Q8BK9_MOUSE	Q3tsg4 mus musculu
21	47	87.0	428	1	FOXDI1_MOUSE	Q8bk9 mus musculu
22	47	87.0	432	2	Q5VYV0_HUMAN	Q64733 mus musculu
23	47	87.0	435	2	Q6GLG1_XENTR	Q5vyv0 homo sapien
24	47	87.0	458	2	Q6P6C2_HUMAN	Q6glg1 xenopus tro
25	47	87.0	506	2	Q3S2W8_ACHDO	Q6p6c2 homo sapien
26	47	87.0	512	1	ZETB8_HUMAN	Q3s2w8 acheta dome
27	47	87.0	521	2	Q7PPS0_ANOGA	Q8nap8 homo sapien
28	47	87.0	529	2	Q4P290_USTMA	Q7pps0 anopheles g
29	47	87.0	578	1	SOB_DROME	Q4p290 ustilago ma
30	47	87.0	708	2	Q9NVRO_HUMAN	Q9vgs7 drosophila
31	47	87.0	708	2	Q8BR71_MOUSE	Q9nvro homo sapien
						Q8br71 mus musculu

32	47	87.0	709	2	Q8CE33_MOUSE	Q8ce33 mus musculu
33	47	87.0	815	1	PYGO_DROME	Q9v9w8 drosophila
34	47	87.0	828	2	Q3UG42_MOUSE	Q3ug42 mus musculu
35	47	87.0	892	2	Q70254_MOUSE	Q70254 mus musculu
36	47	87.0	897	2	Q9VE88_DROME	Q9ve88 drosophila
37	47	87.0	914	2	Q69Z11_MOUSE	Q69z11 mus musculu
38	47	87.0	935	2	Q3UYA4_MOUSE	Q3uya4 mus musculu
39	47	87.0	946	2	Q9VI93_DROME	Q9vi93 drosophila
40	47	87.0	964	1	LBXCO_MOUSE	Q8bx46 mus musculu
41	47	87.0	964	1	LBXCO_RAT	P84551 rattus norv
42	47	87.0	965	1	LBXCO_HUMAN	P84550 homo sapien
43	47	87.0	1065	2	Q4SPK5_TETNG	Q4spk5 tetraodon n
44	47	87.0	1181	2	Q8IR69_DROME	Q8ir69 drosophila
45	47	87.0	1199	2	Q59E49_DROME	Q59e49 drosophila
46	47	87.0	1203	2	Q819J7_DROME	Q819j7 drosophila
47	47	87.0	1248	2	Q9VVF3_DROME	Q9vzf3 drosophila
48	47	87.0	1255	2	Q819J6_DROME	Q819j6 drosophila
49	47	87.0	1298	2	Q5BIA7_DROME	Q5bia7 drosophila
50	47	87.0	1429	2	Q4PAK6_USTMA	Q4pak6 ustilago ma
51	47	87.0	1533	1	PUM_DROME	P25822 drosophila
52	47	87.0	1924	2	Q2KML0_RAT	Q2kml0 rattus norv
53	47	87.0	1925	2	Q2KML1_RAT	Q2kml1 rattus norv
54	47	87.0	1927	2	Q2KMK9_RAT	Q2kmk9 rattus norv
55	47	87.0	1959	2	Q2KMK7_RAT	Q2kmk7 rattus norv
56	47	87.0	1976	2	Q4QHY5_LEIMA	Q4qhy5 leishmania
57	47	87.0	2000	1	CHD3_HUMAN	Q12873 homo sapien
58	47	87.0	2038	1	FSH_DROME	P13709 drosophila
59	47	87.0	2055	2	Q5NCG1_MOUSE	Q5ncg1 mus musculu
60	47	87.0	2145	2	Q9W003_DROME	Q9w003 drosophila
61	47	87.0	3941	2	Q623D7_CAEBR	Q623d7 caenorhabdi
62	46	85.2	52	2	Q5GZW8_ARATH	Q5gzw8 arabidopsis
63	46	85.2	71	2	Q8NI53_HUMAN	Q8ni53 homo sapien
64	46	85.2	112	2	Q3UUS5_MOUSE	Q3uus5 mus musculu
65	46	85.2	117	2	Q3V062_MOUSE	Q3v062 mus musculu
66	46	85.2	130	2	Q9TUC9_CANPA	Q9tuc9 canis famil
67	46	85.2	131	2	Q9TST8_FELCA	Q9ts8 felis silve
68	46	85.2	132	2	Q4TBK8_TETNG	Q4tbk8 tetraodon n
69	46	85.2	143	2	Q52HB4_CHICK	Q52hb4 gallus gall
70	46	85.2	145	2	Q2XYX3_ANNPU	Q2xyx3 annella pu
71	46	85.2	149	2	Q7Q8K9_ANOGA	Q7q8k9 anopheles g
72	46	85.2	150	2	Q9Y4M1_HUMAN	Q9y4m1 homo sapien
73	46	85.2	153	2	Q765Q0_RAT	Q765q0 rattus norv
74	46	85.2	157	2	Q62GX4_BURMA	Q62gx4 burkholderi
75	46	85.2	158	2	Q9MBF7_LILLO	Q9mbf7 lilium long
76	46	85.2	164	2	Q7Q9M3_ANOGA	Q7q9m3 anopheles g
77	46	85.2	164	2	Q8BNH5_MOUSE	Q8bnh5 mus musculu
78	46	85.2	167	2	Q765P9_RAT	Q765p9 rattus norv
79	46	85.2	171	2	Q8H8G6_ORYSA	Q8h8g6 oryza sativ
80	46	85.2	191	2	Q96NI3_HUMAN	Q96ni3 homo sapien
81	46	85.2	199	2	Q4RK77_TETNG	Q4rk77 tetraodon n
82	46	85.2	203	2	Q5SC14_HUMAN	Q5sc14 homo sapien
83	46	85.2	213	2	Q9TUC8_MONDO	Q9tuc8 monodelphis
84	46	85.2	217	1	HAND2_MOUSE	P61296 homo sapien
85	46	85.2	217	1	HAND2_MOUSE	Q61039 m heart- an
86	46	85.2	217	1	HAND2_MOUSE	P61295 rattus norv
87	46	85.2	217	2	Q9EPN2_MOUSE	Q9epn2 mus musculu
88	46	85.2	218	2	Q53G20_HUMAN	Q53g20 homo sapien
89	46	85.2	220	2	Q6IPH7_HUMAN	Q6iph7 homo sapien
90	46	85.2	223	2	Q5ISR0_MACFA	Q5isr0 macaca fasc
91	46	85.2	227	2	Q7Q706_ANOGA	Q7q706 anopheles g
92	46	85.2	228	2	Q5K4L2_9MYRI	Q5k4l2 glomeris ma
93	46	85.2	233	1	ASCL1_RAT	P19359 rattus norv
94	46	85.2	233	2	Q7XTV6_ORYSA	Q7xtv6 oryza sativ
95	46	85.2	236	1	ASCL1_HUMAN	P50553 homo sapien
96	46	85.2	236	2	Q6QDA4_HUMAN	Q6qda4 homo sapien
97	46	85.2	236	2	Q9BX46_HUMAN	Q9bx46 homo sapien
98	46	85.2	244	2	Q6CNB7_KLJLA	Q6cnb7 kluyveromyc
99	46	85.2	247	2	Q9DDF6_PETMA	Q9ddf6 petromyzon
100	46	85.2	248	2	Q9FLDI_ARATH	Q9fldi arabidopsis

ALIGNMENTS

```

RA Ikeno M., Ito S., Ito T., Ito Y., Ito Y., Iwabuchi A., Kamiya K.,
RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.-I., Eun M.-Y.,
RA Yano M., Jiang J., Gojobori T.,
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP003074; BAB40060.1; -; Genomic DNA.
DR EMBL; AP003196; BAB93147.1; -; Genomic DNA.
DR Gramene; Q9ARR7; -.
DR InterPro; IPR009148; Siba.
DR PRINTS; PR01852; SIBAPROTEIN.
KW Hypothetical protein.
SQ SEQUENCE 269 AA; 27747 MW; 861429FCDB1141DF CRC64;

Query Match 92.6%; Score 50; DB 2; Length 269;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFAAAAAAAAAA 13
DB 231 AFAAAAAAAAAA 242
|||||
Q81MU7 DROME PRELIMINARY; PRT; 305 AA.
AC Q81MU7_
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-WAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE CG31422-PA.
GN ORFNames=CG31422, Dmel_CG31422;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

```

```

RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.-I., Eun M.-Y.,
RA Yano M., Jiang J., Gojobori T.,
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AP003074; BAB40060.1; -; Genomic DNA.
DR EMBL; AP003196; BAB93147.1; -; Genomic DNA.
DR Gramene; Q9ARR7; -.
DR InterPro; IPR009148; Siba.
DR PRINTS; PR01852; SIBAPROTEIN.
KW Hypothetical protein.
SQ SEQUENCE 269 AA; 27747 MW; 861429FCDB1141DF CRC64;

Query Match 92.6%; Score 50; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFAAAAAAAAAA 13
DB 51 AFAAAAAAAAAA 62
|||||
Q9ARR7 ORYZA PRELIMINARY; PRT; 269 AA.
AC Q9ARR7_
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 21-FEB-2006, entry version 15.
DE OSJNB0004G10.7 protein (Hypothetical protein B1008C01.41).
GN Name=OSJNB0004G10.7; Synonyms=B1008C01.41;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BAP clade;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=23237276; PubMed=12447438; DOI=10.1038/nature01184;
RA Sasaki T., Matsunoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA Wu J., Niimura Y., Cheng Z., Nigamur Y., Antonio B.A., Kanamori H.,
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,

```

```
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reineck K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sigen-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong Y., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
RL [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celnikier S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bargman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celnikier S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnikier S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX Berkeley Drosophila Genome Project;
RA Celnikier S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
RN [7]
RP NUCLEOTIDE SEQUENCE.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE003750; AAN14029.1; -; Genomic_DNA.
DR FlyBase; FBgn0051422; CG31422.
SQ SEQUENCE 305 AA; 32565 MW; 472F8AA40CFF9953 CRC64;
Query Match 90.7%; Score 49; DB 2; Length 305;
Best Local Similarity 92.3%; Pred. No. 59;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAFAAAAAAAAA 13
```

```
Db 271 AAFAAAAAAAAA 283
RESULT 4
RLAO EIMTE STANDARD; PRT; 314 AA.
ID RLAO_EIMTE
AC Q967Y7;
DT 20-JUN-2002, integrated into UniProtKB/Swiss-Prot.
DT 01-DEC-2001, sequence version 1.
DE 07-FEB-2006, entry version 21.
DE 60S acidic ribosomal protein P0.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5802;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=PAP38;
RA Labbe M., Pery P.;
RT "Characterization of ribosomal phosphoprotein P0 of Eimeria tenella.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Ribosomal protein P0 is the functional equivalent of
CC E.coli protein L10.
CC -!- SUBUNIT: P0 forms a pentameric complex by interaction with dimers
CC of P1 and P2.
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the ribosomal protein L10P family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF353516; AAK38887.1; -; mRNA.
DR InterPro; IPR001813; Ribosomal_60S.
DR InterPro; IPR001790; Ribosomal_L10.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00428; Ribosomal_60S; 1.
DR Pfam; PF00466; Ribosomal_L10; 1.
DR PRINTS; PR00456; RIBOSOMALP2.
KW Phosphorylation; Ribonucleoprotein; Ribosomal protein.
FT CHAIN 1 314 /FTID=PRO_0000154770.
SQ SEQUENCE 314 AA; 33272 MW; 13214EF1694BDAC6 CRC64;
Query Match 90.7%; Score 49; DB 1; Length 314;
Best Local Similarity 92.3%; Pred. No. 60;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAFAAAAAAAAA 13
Db 272 AAFAAAAAAAAA 284
RESULT 5
ID Q5B2U7_EMENI PRELIMINARY; PRT; 399 AA.
AC Q5B2U7;
DT 26-APR-2005, integrated into UniProtKB/TrEMBL.
DT 26-APR-2005, sequence version 1.
DT 07-MAR-2006, entry version 5.
DE Hypothetical protein.
GN ORFNames=AN5133.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=FGSC 4;
RX PubMed=16372000; DOI=10.1038/nature04341;
RA Galagan J.E., Calvo S.E., Cuomo C., Ma L.-J., Wortman J.R.,
RA Batzoglou S., Lee S.-I., Bastuerkmen M., Spevak C.C., Clutterbuck J.,
RA Kapitonov V., Jurka J., Sczzocchio C., Farman M., Butler J.,
```

RA Purcell S., Harris S., Braus G.H., Draht O., Busch S., D'Enfert C.,  
RA Bouchier C., Goldman G.H., Bell-Pedersen D., Griffiths-Jones S.,  
RA Doonan J.H., Yu J., Vlenken K., Pain A., Freitag M., Selker E.U.,  
RA Archer D.B., Penalva M.A., Oakley B.R., Momany M., Tanaka T.,  
RA Kumagai T., Asai K., Machida M., Nierman W.C., Denning D.W.,  
RA Caddick M., Hynes M., Paoletti M., Fischer R., Miller B.L., Dyer P.S.,  
RA Sachs M.S., Osmani S.A., Birren B.W.;  
RA "Sequencing of Aspergillus nidulans and comparative analysis with A.  
RT fumigatus and A. oryzae";  
RL Nature 438:1105-1115(2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AACD01000088; EAA62314.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 399 AA; 43735 MW; 8D25BD74EFA8AD24 CRC64;  
  
Query Match 90.78; Score 49; DB 2; Length 399;  
Best Local Similarity 92.38; Pred. No. 72;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 AAFAAAAA 13  
Db 126 AEFAAAAA 138  
  
RESULT 6  
Q9W4E5 DROME PRELIMINARY; PRT; 422 AA.  
ID Q9W4E5 DROME  
AC Q9W4E5  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 01-MAY-2000, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE CG15465-PA (IP12315p).  
GN ORFNames=CG15465, Dmel CG15465;  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
[1]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Beres P.V., Bertram P.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei E., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merskulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Zheng L.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RG Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleeb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence";  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RT Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleeb J.,  
RA Park S., Wan K., Yu C., Celniker S.;  
RA Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AE003433; AAP46008.1; -; Genomic\_DNA.  
DR EMBL; BT022700; AAY55116.1; -; mRNA.  
DR FlyBase; FBgn0029746; CGI5465.  
SQ SEQUENCE 422 AA; 46793 MW; 3DF7F9EAF6A3DD1E CRC64;  
  
Query Match 90.78; Score 49; DB 2; Length 422;

```
Best Local Similarity 92.3%; Pred. No. 75;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cc 1 AAAAAAAAAAAAAA 13
Cc 65 AAAAAAAAAAAAAA 77

RESULT 7
FOXDL_MOUSE
ID FOXDL_MOUSE STANDARD; PRT; 456 AA.
AC Q61345;
DT 15-JUL-1998, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1996, sequence version 1.
DE Forkhead box protein D1 (Forkhead-related protein FKHL8) (Forkhead-
DE related transcription factor 4) (FREAC-4) (Brain factor-2) (BFH-BF-2).
GN Name=Foxd1; Synonyms=Fkhl8, Freac4, Fhbf2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA].
RC STRAIN=BALB/c; TISSUE=Embryo;
RX MEDLINE=95114592; PubMed=7815060;
RA Hatini V., Tao W., Lai E.;
RT "Expression of winged helix genes, BF-1 and BF-2, define adjacent
RT domains within the developing forebrain and retina.";
RL J. Neurobiol. 25:1293-1309(1994).
RN [2]
RP FUNCTION.
RX PubMed=15509772; DOI=10.1242/dev.01431;
RA Herrera E., Marcus R., Li S., Williams S.E., Erskine L., Lai E.,
RA Mason C.;
RT "Foxd1 is required for proper formation of the optic chiasm.";
RL Development 131:5727-5739(2004).
RN [3]
RP FUNCTION.
RX PubMed=15634693; DOI=10.1242/dev.01604;
RA Levinson R.S., Batorina E., Choi C., Vorontchikhina M.,
RA Kitajewski J., Mendelsohn C.L.;
RT "Foxd1-dependent signals control cellularity in the renal capsule, a
RT structure required for normal renal development.";
RL Development 132:529-539(2005).
CC -!- FUNCTION: Transcription factor required for formation of
CC positional identity in the developing retina, regionalization of
CC the optic chiasm and morphogenesis of the kidney.
CC -!- SUBCELLULAR LOCATION: Nucleus (Potential).
CC -!- TISSUE SPECIFICITY: Predominantly expressed in the CNS and
CC temporal half of the retina. Also expressed in the condensed head
CC mesenchyme, metanephric blastema of the developing kidney, cortex
CC of the adrenal gland, condensed mesenchyme at the base of the
CC follicles of vibrissae and cartilage perichondrium of the
CC developing vertebrae.
CC -!- DEVELOPMENTAL STAGE: At E9.5 embryos, expressed in a limited
CC region of the neuroepithelium and also in the temporal half of the
CC primary optic cup and the optic stalk. At E10.5, seen in the
CC hypothalamus, temporal half of the optic stalk, and temporal
CC hemiretina. At E12.5 and E13.5 a high expression is seen in
CC regions of condensed mesenchyme of the head, and as
CC neuroepithelial cells begin to differentiate and migrate outward
CC from the ventricular zone, expression declines markedly. By E16.5
CC levels are diminished and restricted to unfused pockets along the
CC exhausted ventricular zone.
CC -!- MISCELLANEOUS: Mice lacking Foxd1 show disrupted cell identity in
CC the ventrotemporal area of the retina and aberrant morphogenesis
CC of the optic chiasm. Their kidneys remain fused, have a
CC disorganised ureteric tree and fail to ascend to a lumbar
CC position.
CC -!- SIMILARITY: Contains 1 fork-head DNA-binding domain.
CC -!- CAUTION: Was originally (Ref.1) assigned to be BF-2 (FOXG1A).
```

```
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; L38607; AAC42042.1; -; mRNA.
HSSP; Q63245; 2HDC.
SMR; Q61345; 129-225.
DR TRANSFAC; T02293; -.
DR Ensembl; ENSMUSG0000050789; Mus musculus.
DR MGI; MGI:1347463; Foxd1.
DR GO; GO:0003677; F-DNA binding; IDA.
DR GO; GO:0007411; P-axon guidance; IMP.
DR GO; GO:0008357; P-regulation of transcription from RNA polyme. .; IDA.
DR InterPro; IPR001766; TF_Fork_head.
DR InterPro; IPR011991; Wing_hlx_DNA_bd.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW Developmental protein; DNA-Binding; Nuclear protein; Transcription;
KW Transcription regulation.
FT CHAIN 1 456 Forkhead box protein D1.
FT /FTID=PRO_0000091812.
FT DNA_BIND 130 224 Fork-head.
FT COMBIAS 29 32 Poly-Glu.
FT COMBIAS 33 36 Poly-Asp.
FT COMBIAS 38 46 Poly-Gly.
FT COMBIAS 55 58 Poly-Arg.
FT COMBIAS 73 76 Poly-Asp.
FT COMBIAS 261 267 Poly-Pro.
FT COMBIAS 293 302 Poly-Ala.
FT COMBIAS 308 319 Poly-Pro.
FT COMBIAS 395 400 Poly-Gly.
FT COMBIAS 420 425 Poly-Ala.
SQ SEQUENCE 456 AA; 45430 MW; 43D93F89BDBFCCC7 CRC64;

Query Match 90.7%; Score 49; DB 1; Length 456;
Best Local Similarity 92.3%; Pred. No. 80;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 290 ALFAAAAAAAAAAAAA 302

RESULT 8
Q3UQW8_MOUSE PRELIMINARY; PRT; 456 AA.
ID Q3UQW8_MOUSE
AC Q3UQW8;
DT 11-OCT-2005, integrated into UniProtKB/TREMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE 12 days embryo eyeball cDNA, RIKEN full-length enriched library,
DE clone:023000910 product:forkhead box D1, full insert sequence (16
DE days embryo head cDNA, RIKEN full-length enriched library,
DE clone:4122401E09 product:forkhead box D1, full insert sequence).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;
RX MEDLINE=99779253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
```



RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard D., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Altshuler S., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgi-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hilt D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J.C., Reid J.P., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.B., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki K., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano N., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RA RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core team) and the FANTOM Consortium;  
 RT "Antisense transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Head, and Eyeball;  
 RX MEDLINE=22354685; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedziarski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrowsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shinada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,

RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;  
 RX MEDLINE=251085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bono M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hatada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball, and Head;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Horii F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 RL !- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC EMBL; AK142034; BAE24920.1; -; mRNA.  
 CC EMBL; AK132390; BAE21140.1; -; mRNA.

```
DR GO: 0003677; F:DNA binding; IDA.
DR GO: 0007411; P:axon guidance; IMP.
DR GO: 0006357; P:regulation of transcription from RNA polyme...; IDA.
DR InterPro: IPR001766; TF_Fork head.
DR InterPro: IPR00217; Tubulin.
DR Pfam: PF00250; Fork head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR ProDom: PD000425; TF_Fork_head; 1.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK HEAD 1; 1.
DR PROSITE: PS00658; FORK HEAD 2; 1.
DR PROSITE: PS00039; FORK HEAD 3; 1.
DR PROSITE: PS00227; TUBULIN; UNKNOWN 1.
KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.
SQ SEQUENCE 456 AA; 45371 MW; 3153172991DC9A4 CRC64;

Query Match          90.7%; Score 49; DB 2; Length 456;
Best Local Similarity 92.3%; Pred. No. 80;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAFAAAAA 13
Db 290 ALFAAAAA 302

RESULT 9
FOXDI1 HUMAN STANDARD; PRT; 465 AA.
AC Q16676; Q12949;
DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.
DT 01-NOV-1997, sequence version 1.
DT 07-FEB-2006, entry version 33.
DE Forkhead box protein D1 (Forkhead-related protein FKHL8) (Forkhead-
DE related transcription factor 4) (FREAC-4).
GN Name=FOXDI; Synonyms=FKHL8, FREAC4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA / MRNA].
RX MEDLINE=96355467; PubMed=8702877; DOI=10.1074/jbc.271.35.21094;
RA Ernstsson S., Pierrou S., Hulander M., Cedersberg A., Hellqvist M.,
RA Carlsson P., Enerbaeck S.;
RT "Characterization of the human forkhead gene FREAC-4. Evidence for
RT regulation by Wilms' tumor suppressor gene (WT-1) and p53.";
RL J. Biol. Chem. 271:21094-21099(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE [MRNA] OF 120-225.
RX MEDLINE=95045392; PubMed=7957066;
RA Pierrou S., Hellqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;
RT "Cloning and characterization of seven human forkhead proteins:
RT binding site specificity and DNA bending.";
RL EMBO J. 13:5002-5012(1994).
CC -!- FUNCTION: Transcription factor required for formation of
CC positional identity in the developing retina, regionalization of
CC the optic chiasm and morphogenesis of the kidney (by similarity).
CC -!- SUBCELLULAR LOCATION: Nucleus.
CC -!- SIMILARITY: Contains 1 fork-head DNA-binding domain.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL: U59831; AAC50660.1; -; Genomic_DNA.
DR EMBL: U59831; AAC50661.1; -; mRNA.
DR EMBL: U13222; AAA92039.1; -; mRNA.
DR PIR: G02738; G02738.
DR PIR: S51627; S51627.
DR HSSP: Q63245; 2HDC.
DR SMR: Q16676; 124-220.
DR TRANSFAC: T02472; -.
DR HGNC: HGNC:3802; FOXD1.
```

```
DR MIM: 601091; gene.
DR GO: 0003700; F:transcription factor activity; TAS.
DR InterPro: IPR001766; TF_Fork head.
DR InterPro: IPR011991; Wing hlx_DNA_bd.
DR Pfam: PF00250; Fork head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR ProDom: PD000425; TF_Fork_head; 1.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00657; FORK HEAD 1; 1.
DR PROSITE: PS00658; FORK HEAD 2; 1.
DR PROSITE: PS00039; FORK HEAD 3; 1.
KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.
FT CHAIN 1 465 /FTId=PRO_0000091811.
FT DNA_BIND 124 215 Fork-head.
FT COMBIAS 26 34 Poly-Glu.
FT COMBIAS 39 43 Poly-Gly.
FT COMBIAS 52 57 Poly-Arg.
FT COMBIAS 69 72 Poly-Glu.
FT COMBIAS 73 76 Poly-Asp.
FT COMBIAS 97 113 Poly-Gly.
FT COMBIAS 231 234 Poly-Ala.
FT COMBIAS 252 256 Poly-Ala.
FT COMBIAS 259 266 Poly-Pro.
FT COMBIAS 293 303 Poly-Ala.
FT COMBIAS 309 315 Poly-Pro.
FT COMBIAS 375 390 Poly-Ala.
FT COMBIAS 428 434 Poly-Ala.
SQ SEQUENCE 465 AA; 46140 MW; D3E7854909CCBFAE CRC64;

Query Match          90.7%; Score 49; DB 1; Length 465;
Best Local Similarity 92.3%; Pred. No. 81;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RAFAAAAA 13
Db 290 ALFAAAAA 302

RESULT 10
Q6Z8T9 ORYZA PRELIMINARY; PRT; 167 AA.
ID Q6Z8T9 ORYZA PRELIMINARY; PRT; 167 AA.
AC Q6Z8T9;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 14.
DE Zinc finger protein family-like.
DE Name=P06861.17;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- DOMAIN: The RING-type zinc finger domain is essential for
CC ubiquitin ligase activity. It coordinates an additional third zinc
CC ion (By similarity).
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
DR EMBL: AP004762; BAD10011.1; -; Genomic_DNA.
DR Gramene: Q6Z8T9; -.
DR GO: 0005634; C:nucleus; IEA.
DR GO: 0000151; C:ubiquitin ligase complex; IEA.
DR GO: 00046872; F:metal ion binding; IEA.
DR GO: 0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO: 0008270; F:zinc ion binding; IEA.
DR GO: 0016567; P:protein ubiquitination; IEA.
```

```
DR InterPro; IPR011016; RINGV.
DR InterPro; IPR001841; Znf_RING.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00744; RINGV; 1.
DR PROSITE; PSS0089; ZF_RING_2; 1. Zinc-finger.
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.
SQ SEQUENCE 167 AA; 17664 MW; 4729ACE00A901AE9 CRC64;

Query Match      88.9%; Score 48; DB 2; Length 167;
Best Local Similarity 92.3%; Pred. No. 50;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
   |||||
DB 56 AATAAAAA 68

RESULT 11
ID Q4PEI8 USTMA PRELIMINARY; PRT; 361 AA.
AC Q4PEI8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DE Hypothetical protein.
DE Hypothetical protein.
GN ORFNames=UM01475.1;
OS Ustilago maydis 521.
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=237631;
RN [1]
NP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=521;
RA Birren B.W., Nubaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H.M., Ambrester J., Bachantsang J., Baldwin J., Barry A.,
RA Bayl T., Blitshetyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Bouhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S.E., Camarata J., Campo K., Chang J., Cheshatsang Y., Corum B.,
RA Citroen M., Collymore A., Considine T., Cook A., Cooke P., Corum B.,
RA Cuomo C., David R., Dawoe T., Degray S., Dodge S., Dooley K.,
RA Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,
RA Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J.E., Gearin G., Gnerre S.,
RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseles M., Karlsson E.,
RA Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-Toh K., Liu X., Lokysang T., Lokysang Y., Lucien O.,
RA Lui A., Ma L.-J., Mabbitt R., MacDonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., McGehee T., Meldrum J., Meneus L.,
RA Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA Moraes J., Mulrain L., Nelson G., Naylor J., News C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C.B., Nizzari M., Norbu C.,
RA Norbu N., O'Donnell P., Okaawa O., O'Leary S., Omotohso B.,
RA O'Neill K., O'Sullivan S., Parker S., Perrin D., Phunkhang P., Pigani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-Thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V.S., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.S.;
RT "The genome sequence of Ustilago maydis.";
```

```
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AACP01000053; EAK82242.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 361 AA; 39615 MW; 038F774408202041 CRC64;

Query Match      88.9%; Score 48; DB 2; Length 361;
Best Local Similarity 92.3%; Pred. No. 90;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
   |||||
DB 133 AATAAAAA 145

RESULT 12
ID ZBTB8 MOUSE STANDARD; PRT; 484 AA.
AC Q8CII0;
DT 26-APR-2004, integrated into UniProtKB/Swiss-Prot.
DT 01-MAR-2003, sequence version 1.
DT 07-MAR-2006, entry version 22.
DE Zinc finger and BTB domain-containing protein 8.
GN Name=Zbtb8;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
NP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=FVB/N;
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Hellon E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL CC -! FUNCTION: May be involved in transcriptional regulation.
CC -! SUBCELLULAR LOCATION: Nucleus (Potential).
CC -! SIMILARITY: Contains 1 BTB (POZ) domain.
CC -! SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC023839; AAR23839.1; -; mRNA.
DR Ensembl; ENSMUSG00000048485; Mus musculus.
DR MGI; MGI:2387181; Zbtb8.
DR InterPro; IPR000210; BTB.
DR InterPro; IPR013049; BTB_POZ.
DR InterPro; IPR007087; Znf_C2H2.
```

DR Pfam: PF00651; BTB; 1.  
DR Pfam: PF00096; zf-C2H2; 2.  
DR ProDom: PD000003; Znf\_C2H2; 1.  
DR SMART: SM00225; BTB; 1.  
DR SMART: SM00355; Znf\_C2H2; 2.  
DR PROSITE: PS00097; BTB; 1.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_2; 2.  
DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
KW DNA-binding; Metal-binding; Nuclear protein; Repeat; Transcription;  
Transcription regulation; Zinc; Zinc-finger.  
FT CHAIN 1 484  
FT Zinc finger and BTB domain-containing  
FT protein 8.  
FT /FTID=PRO\_0000047722.  
FT  
FT DOMAIN 24 92  
FT ZN\_FING 331 353  
FT ZN\_FING 359 382  
FT COMPBIA5 132 149  
FT Ala-rich.  
SQ SEQUENCE 484 AA; 53387 MW; B4PF082555B0CA1A CRC64;  
Query Match 88.9%; Score 48; DB 1; Length 484;  
Best Local Similarity 92.3%; Pred. No. 1.1e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAFRAAAAAAAAAA 13  
Dy 136 AAAAAAAAAAAAAA 148  
|| |||||  
|| |||||  
RESULT 13  
Q3US18 MOUSE PRELIMINARY; PRT; 484 AA.  
AC Q3US18;  
DT 11-OCT-2005, sequence version 1.  
DT 21-FEB-2006, entry version 7.  
DE 16 days embryo head cDNA, RIKEN full-length enriched library,  
DE clone: C130068L16 product: zinc finger and BTB domain containing 8, full  
DE insert sequence (0 day neonate cerebellum cDNA, RIKEN full-length  
DE enriched library, clone: C230068H07 product: zinc finger and BTB domain  
DE containing 8, full insert sequence) (14, 17 days embryo head cDNA,  
DE RIKEN full-length enriched library, clone: 3222401B20 product: zinc  
DE finger and BTB domain containing 8, full insert sequence).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RC MEDLINE=92279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RC PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Belsel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,

RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera F.M., Marchionni L.,  
RA Matsuoka H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugitani K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RC MEDLINE=16141073; DOI=10.1126/science.1112009;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
[4]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RX Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nika I., Oatso N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.B., Cousins S.,  
RA Dalla E., Dragni T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kugawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
[5]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RX Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

DR InterPro: IPR007087; Znf\_C2H2.  
DR Pfam: PF00651; BTE; 1.  
DR Pfam: PF00096; zf-C2H2; 2.  
DR ProDom: PD000003; Znf\_C2H2; 1.  
DR SMART: SM00225; BTE; 1.  
DR SMART: SM00355; Znf\_C2H2; 2.  
DR PROSITE: PS50097; BTE; 1.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 2.  
KW Metal-binding; Nuclear protein; Repeat; Zinc; Zinc-finger.  
SQ SEQUENCE 484 AA; 53414 MW; ACA81D5402A2E7F4 CRC64;

Query Match 88.9%; Score 48; DB 2; Length 484;  
Best Local Similarity 92.13%; Pred. No. 1.1e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 1 AAAAAAAAAAAAA 13  
|||  
DB 136 AAAAAAAAAAAAA 148

RESULT 14  
Q9VTW5 DROME  
ID Q9VTW5 DROME PRELIMINARY; PRT; 640 AA.  
AC Q9VTW5;  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2002, sequence version 2.  
DT 21-FEB-2006, entry version 39.  
DE CG32105-PB (R570810p).  
DR ORFNames=CG32105, Dmel\_CG32105;  
GN Drosophila melanogaster (Fruit fly).  
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP UNCLEOTTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celinick S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,  
RA Foeller C., Gabriellani A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegman C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E.E., Sridling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissensbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celnikier S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
 RA Pacleeb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celnikier S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomics perspective.";  
 RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettecourt B.R., Celnikier S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RX Berkeley *Drosophila* Genome Project;  
 RA Celnikier S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
 RA Hoskins R., Stapleton M., Pacleeb J., Park S., Svirskas R., Smith E.,  
 RA Yu C., Rubin G.;  
 RT "*Drosophila melanogaster* release 4 sequence.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Flybase;  
 RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN-Berkeley;  
 RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleeb J.,  
 RA Park S., Wan K., Yu C., Celnikier S.;  
 RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AE003541; AAF49930.2; -; Genomic\_DNA.  
 DR EMBL; BT003467; AAO39470.1; -; mRNA.  
 DR HSSP; P50480; 1BW5.  
 DR IntAct; Q9VTW5; -.  
 DR FlyBase; FBgn0052105; CG32105.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR012287; Homeodomain-rel.  
 DR InterPro; IPR007107; LIM homeo.  
 DR InterPro; IPR001781; LIM Zn bd.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR Pfam; PF00046; Homeobox; 1.

DR Pfam; PF00412; LIM; 2.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR ProDom; PD000094; LIM; 2.  
 DR SMART; SM00389; HOK; 1.  
 DR SMART; SM00132; LIM; 2.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS00071; HOMEBOX\_2; 1.  
 DR PROSITE; PS00478; LIM DOMAIN\_1; 2.  
 DR PROSITE; PS00023; LIM DOMAIN\_2; 2.  
 DR PROSITE; PS01359; ZF\_PHD\_1; UNKNOWN 1.  
 KW DNA-binding; Developmental protein; Homeobox; LIM domain;  
 KW Metal-binding; Nuclear protein; Transcription;  
 KW Transcription regulation; Zinc.  
 SQ SEQUENCE 640 AA; 67760 MW; A1A10F826018C98E CRC64;  
 Query Match 88.9%; Score 48; DB 2; Length 640;  
 Best Local Similarity 92.3%; Pred.No. 1.4e+02;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAFAAAAAAAAAA 13  
 DB 189 AAIAAAAAAAAAA 201  
 RESULT 15  
 Q9FYS1\_HORVND PRELIMINARY; PRT; 339 AA.  
 ID Q9FYS1\_HORVND PRELIMINARY; PRT; 339 AA.  
 AC Q9FYS1;  
 DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 17.  
 DE Glutamy1-tRNA reductase (Fragment).  
 GN Name=hema2;  
 OS Hordeum vulgare var. distichum (Two-rowed barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
 OC Poaceae; Triticeae; Hordeum.  
 OX NCBI\_TaxID=112509;  
 RN [1]\_TaxID=112509;  
 RP NUCLEOTIDE SEQUENCE.  
 RA Hansson M.;  
 RT "Basic characterization of two barley hema promoter regions reveals  
 RT stem-loop structures and suggests a regulatory role in Poaceae  
 RT tetrapyrrole biosynthesis.";  
 RL Plant Physiol. Biochem. 39:155-160(2001).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AF294753; AAG02480.1; -; Genomic\_DNA.  
 DR HSSP; Q9UXR8; 1GPU.  
 DR Gramene; Q9FYS1; -.  
 DR GO; GO:0005737; C:cytoplasm; IEA.  
 DR GO; GO:0008883; F:glutamy1-tRNA reductase activity; IEA.  
 DR GO; GO:0004764; F:shikimate 5-dehydrogenase activity; IEA.  
 DR GO; GO:0016089; P:aromatic amino acid family biosynthesis, sh. .; IEA.  
 DR GO; GO:0006779; P:porphyrin biosynthesis; IEA.  
 DR InterPro; IPR000343; GlutR.  
 DR InterPro; IPR006151; Shikimate\_DH.  
 DR Pfam; PF05201; GlutR\_N; 1.  
 DR Pfam; PF01488; Shikimate\_DH; 1.  
 DR PROSITE; PS00747; GLUTR; 1.  
 FT NON\_TER 339  
 SQ SEQUENCE 339 AA; 36103 MW; 7499C62E55638735 CRC64;  
 Query Match 87.0%; Score 47; DB 2; Length 339;  
 Best Local Similarity 91.7%; Pred.No. 1.2e+02;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAFAAAAAAAAAA 12  
 DB 9 AAFAAAAAAAAAA 20

```

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Smerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Mouse;
RG NIH MGC Project;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC052076; AAH52076.1; -; mRNA.
CC Ensembl; ENSMUSG0000042650; Mus musculus.
DR MGI; MGI:2144489; AW050020.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR InterPro; IPR005123; 2OG-FeII_Oase.
DR Pfam; PF03171; 2OG-FeII_Oxy; 1.
DR Hypothetical protein; Iron; Oxidoreductase.
KW Hypothetical protein; Iron; Oxidoreductase.
SQ SEQUENCE 395 AA; 44410 MW; 4CE25B4FB5B04DEB CRC64;

Query Match 87.0%; Score 47; DB 2; Length 395;
Best Local Similarity 92.3%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
DB 36 AAVAAAAA 48

RESULT 18
Q8BKCI_MOUSE PRELIMINARY; PRT; 395 AA.
AC Q8BKCI_MOUSE
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 22.
DE 0 day neonate eyeball cDNA, RIKEN full-length enriched library,
DE clone: E130207K11 product: hypothetical Alanine-rich region/Type 1
DE antifreeze protein/2OG-Fe(II) oxygenase superfamily containing
DE protein, full insert sequence (2 cells egg cDNA, RIKEN full-length
DE enriched library, clone: B020011005 product: Hypothetical alanine-rich
DE region/type 1 antifreeze protein/2OG-Fe, full insert sequence) (Novel
DE protein).
DE Name=AW050020; ORFNames=RP23-331I21.1-001;
DE Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RL "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells S., Kodzius R., Shimokawa K.,
RA Bajic V.B., Bremner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilmshurst L.G., Aldinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,

```



RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dairymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelson J., Kitamura H.,  
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nelson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reid J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shinada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563 (2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome";  
RL Science 309:1564-1566 (2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaio I., Oshino M., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs";

RL Nature 420:563-573 (2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaio I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690 (2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630 (2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuo S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771 (2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE  
RC STRAIN=C57BL/6J; TISSUE=Egg;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.



```

RN RA NUCLEOTIDE SEQUENCE.
RP Tracey A.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK053695; BAC35478.1; -; mRNA.
DR EMBL; AK163294; BAC37281.1; -; mRNA.
DR EMBL; AL596386; CA135332.1; -; Genomic DNA.
DR Ensembl; ENSMUSG00000042650; Mus musculus.
DR MGI; MGI:2144489; AW050020.
DR GO; GO:000506; F:iron ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR InterPro; IPR005123; 2OG-FeII Oase.
DR Pfam; PF03171; 2OG-FeII Oxy; I
KW Hypothetical protein; iron; Oxidoreductase.
SQ SEQUENCE 395 AA; 4441 MW; 2822BF4BFFFE7EF CRC64;

Query Match 87.0%; Score 47; DB 2; Length 395;
Best Local Similarity 92.3%; Pred. No. 1.3e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAFAAAAA 13
Db 36 AAVAAAAA 48

RESULT 19
Q3TSG4 MOUSE PRELIMINARY; PRT; 395 AA.
AC Q3TSG4;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE In vitro fertilized eggs cDNA, RIKEN full-length enriched library,
DE clone:7420402E17 product:Hypothetical alanine-rich region/type I
DE antifreeze protein/2OG-Fe, full insert sequence.
GN Name=AW050020;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama K., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Allen J.E.,
RA Davis M.J., Wilming L.G., Ainslis V., Allen J.E.,
RA Ambal-Impombato A., Apweiler R., Attalini R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.F., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humnick L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,

```

```

RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sanderlin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wallestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
[3]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
[4]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nitaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Guscinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sanderlin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wallestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[5]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX MEDLINE=251085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

```

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 465:685-690(2001).  
RN [16]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [17]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuana T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura K., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [18]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK162072; BA636711.1; -; mRNA.  
DR MGI; MGI:2144489; AW050020.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR InterPro; IPR005123; 2OG-FeII\_Oase.  
DR Pfam; PF03171; 2OG-FeII\_Oxy; I.  
KW Hypothetical protein.  
SQ SEQUENCE 395 AA; 44494 MW; 1F3B93E0E71AF27D CRC64;  
Query Match 87.0%; Score 47; DB 2; Length 395;  
Best local Similarity 92.3%; Pred. No. 1.3e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 AAFAAAAA 13  
Db 36 AAAAAA 48  
RESULT 20  
Q8BK9 MOUSE PRELIMINARY; PRT; 395 AA.  
ID Q8BK9 MOUSE  
AC Q8BK9;

DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-MAR-2006, entry version 24.  
DE 0 day neonate eyeball cDNA, RIKEN full-length enriched library.  
DE clone:EI3021021 product:hypothetical Alanine-rich region/Type I  
DE antifreeze protein/2OG-Fe(II) oxygenase superfamily containing  
DE protein, full insert sequence.  
DE Name:AW050020;  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Clowe K.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Guscinich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuura H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakamura M., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.P., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Suglura K., Sultana R., Takenaka Y., Taki K.,  
RA Tannoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yang K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Yamashita T., Zdobych R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Finkbeiner S., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 303:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Vagi K., Tonaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hilli D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimond S., Gustincich S., Hirokawa N., Jackson I. J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Matsushima T., Numata K., Okido T., Pavan W.J., Pertea G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kigawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
 RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,

RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multiplexed sequencer";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Sano H.,  
 RA Saito R., Saitoh K., Sakai C., Sakai K., Sakazume N., Sato H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Takawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,  
 RA Tonaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AK053700; BAC35481.1; -; mRNA.  
 DR Ensembl; ENSMUSG00000042650; Mus musculus.  
 DR MGI; MGI:2144489; AW050020.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 395 AA; 44415 MW; 6D22C36B1A99EDA8 CRC64;  
 Query Match 87.0%; Score 47; DB 2; Length 395;  
 Best Local Similarity 92.3%; Pred. No. 1.3e+02;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAAAAAA 13  
 Db 36 AAAAAAAAAAAAAA 48  
 RESULT 21  
 FOXB2\_MOUSE  
 ID FOXB2\_MOUSE STANDARD; PRT; 428 AA.  
 AC Q64733;  
 DT 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1997, sequence version 1.  
 DT 07-FEB-2006, entry version 31.  
 DE Forkhead box protein B2 (Transcription factor FKX-4).  
 GN Name=Foxb2; Synonyms=Fkh4;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Mus.  
 OC NCBI\_TaxId=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE  
 RX MEDLINE=97014266; PubMed=8861101; DOI=10.1016/0925-4773(96)00507-2;  
 RA Kaestner K.H., Schuetz G., Monaghan A.P.;  
 RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains  
 RT in the central nervous system";  
 RL Mech. Dev. 55:221-230(1996).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE OF 4-114.  
 RC STRAIN=129;  
 RX MEDLINE=93361500; PubMed=7689224;  
 RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H., Monaghan A.P.,  
 RA Schuetz G.;  
 RT "Six members of the mouse forkhead gene family are developmentally  
 RT regulated";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
 CC -!- SUBCELLULAR LOCATION: Nucleus.  
 CC -!- DEVELOPMENTAL STAGE: Expressed during embryogenesis.

```
CC CC -!- SIMILARITY: Contains 1 fork-head DNA-binding domain.
CC CC -----
CC CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC CC Distributed under the Creative Commons Attribution-NoDerivs License
CC CC -----
CC CC EMBL; X92591; CA63335.1; -; mRNA.
CC CC EMBL; X71942; CA50744.1; -; Genomic_DNA.
CC CC FIR; D47746; D47746.
CC CC HSSP; Q63245; 2HDC.
CC CC TRANSFAC; T02442; -.
CC CC Ensembl; ENSMUSG0000056829; Mus musculus.
CC CC MGI; MGI:1347468; Foxb2.
CC CC InterPro; IPR001766; TF_Fork_head.
CC CC InterPro; IPR011991; Wing_hlx_DNA_bd.
CC CC Pfam; PF00250; Fork_head; 1.
CC CC PRINTS; PR00053; FORKHEAD.
CC CC ProDom; PD000425; TF_Fork_head; 1.
CC CC SMART; SM00339; FH; 1.
CC CC PROSITE; PS00657; FORK_HEAD_1; 1.
CC CC PROSITE; PS00658; FORK_HEAD_2; 1.
CC CC PROSITE; PS50039; FORK_HEAD_3; 1.
CC CC DNA-binding; Nuclear protein; Transcription; Transcription regulation.
CC CC KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.
CC CC FT CHAIN 1 428 Forkhead box protein B2.
CC CC /FTID=PRO_0000091805.
CC CC FT DNA BIND 12 103 Fork-head.
CC CC FT COMPBIAS 139 153 Poly-His.
CC CC FT COMPBIAS 156 162 Poly-His.
CC CC FT COMPBIAS 163 172 Poly-Pro.
CC CC FT COMPBIAS 217 231 Poly-Ala.
CC CC FT COMPBIAS 249 258 Poly-Ala.
CC CC FT COMPBIAS 321 330 Poly-Ala.
CC CC FT COMPBIAS 396 399 Poly-Ala.
CC CC SQ SEQUENCE 428 AA; 45170 MW; DB8A8EFD1E94AB10 CRC64;

Query Match 87.0%; Score 47; DB 1; Length 428;
Best Local Similarity 92.3%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAAAAA 13
Db 218 AAAAAAAAAAAAAA 230

RESULT 22
Q5VYV0 HUMAN PRELIMINARY; PRT; 432 AA.
AC Q5VYV0;
DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 07-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE OTHUMP0000021510.
GN ORFNames=RP11-159H20.4-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kimberley A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -----
CC CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC CC Distributed under the Creative Commons Attribution-NoDerivs License
CC CC -----
CC CC EMBL; AL353637; CAH70683.1; -; Genomic_DNA.
CC CC GO; GO:0005634; C:nucleus; IEA.
CC CC GO; GO:0003700; P:transcription factor activity; IEA.
CC CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
CC CC GO; GO:0006350; P:transcription; IEA.
CC CC InterPro; IPR001766; TF_Fork_head.
CC CC InterPro; IPR011991; Wing_hlx_DNA_bd.
```

```
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR ProDom; PD000425; TF_Fork_head; 1.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; UNKNOWN_1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.
SQ SEQUENCE 432 AA; 45581 MW; 9A832ACDD9A765EF CRC64;

Query Match 87.0%; Score 47; DB 2; Length 432;
Best Local Similarity 92.3%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAAAAA 13
Db 221 AAAAAAAAAAAAAA 233

RESULT 23
Q6GLG1 XENTR PRELIMINARY; PRT; 435 AA.
AC Q6GLG1;
DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 16.
DE SWI/SNF related, matrix associated, actin dependent regulator of
DE chromatin, subfamily e, member 1.
GN Name=smarcd1-prov;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toehiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC CC Distributed under the Creative Commons Attribution-NoDerivs License
CC CC -----
CC CC EMBL; BC074534; AAH74534.1; -; mRNA.
CC CC Ensembl; ENSXETG00000021355; Xenopus tropicalis.
CC CC GO; GO:0003677; P:DNA binding; IEA.
CC CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
CC CC InterPro; IPR000910; HMG_12_box.
CC CC Pfam; PF00505; HMG_box; 1.
```

```
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS50118; HMG_BOX_2; 1. 6P227A5D387A9CA2 CRC64;
SQ SEQUENCE 435 AA; 48441 MW; 62227A5D387A9CA2 CRC64;

Query Match 87.0%; Score 47; DB 2; Length 435;
Best Local Similarity 92.3%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
  |||
Db 318 AAVAAAAAAAAA 330
  |||

RESULT 24
Q6P6C2 HUMAN
ID Q6P6C2 HUMAN PRELIMINARY; PRT; 458 AA.
AC Q6P6C2;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2006, entry version 16.
DE Hypothetical protein OFOX.
GN Name=OFOX;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klapper S.L., Collins F.S., Wagner L., Shenmen J.C., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RG NIH MGC Project;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; BC062339; AAG62339.1; -; mRNA.
DR Ensemble; ENSG0000091542; Homo sapiens.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR InterPro; IPR005123; 2OG-Ferri_Oase.
DR Pfam; PF03171; 2OG-Ferri_Oxy; 1.
DR Hypothetical protein; Ifon; Oxidoreductase.
KW SEQUENCE 458 AA; 51409 MW; 88ACCD9992B5AEB CRC64;

Query Match 87.0%; Score 47; DB 2; Length 458;
Best Local Similarity 92.3%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DR SMART; SM00398; HMG; 1.
DR PROSITE; PS50118; HMG_BOX_2; 1. 6P227A5D387A9CA2 CRC64;
SQ SEQUENCE 435 AA; 48441 MW; 62227A5D387A9CA2 CRC64;

Query Match 87.0%; Score 47; DB 2; Length 506;
Best Local Similarity 92.3%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAA 13
  |||
Db 218 AAVAAAAAAAAA 230
  |||

RESULT 26
ZBTB8 HUMAN
ID ZBTB8 HUMAN STANDARD; PRT; 512 AA.
AC Q8NAP8; Q5VKR5;
DT 26-APR-2004, integrated into UniProtKB/Swiss-Prot.
DT 01-OCT-2002, sequence version 1.
DT 07-MAR-2006, entry version 25.
DE Zinc finger and BTB domain-containing protein 8.
GN Name=ZBTB8;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Brain;
RX PubMed=14702039; DOI=10.1038/ng1285;
```

Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Negai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto K., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,  
RA Niimiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Yuashino K., Yuuki H., Oshima A., Sasaki N., Rotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs";  
RL Nat. Genet. 36:40-45(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RG Human chromosome 1 international sequencing consortium;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: May be involved in transcriptional regulation.  
CC -!- SUBCELLULAR LOCATION: Nucleus (Potential).  
CC -!- SIMILARITY: Contains 1 BTB (POZ) domain.  
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AK092326; BAC03863.1; -; mRNA.  
DR EMBL; AL356986; CAH73365.1; -; Genomic DNA.  
DR EMBL; AL033529; CAH73365.1; JOINED; Genomic DNA.  
DR EMBL; AL033529; CAI20022.1; -; Genomic DNA.  
DR EMBL; AL356986; CAI20022.1; JOINED; Genomic DNA.  
DR EMBL; ENSG00000185718; Homo sapiens.  
DR HGNC; HGNC:24172; ZBTB8.  
DR InterPro; IPR000210; BTB.  
DR InterPro; IPR013069; BTB\_POZ.  
DR InterPro; IPR007087; ZNF\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf-C2H2; 2.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; ZNF\_C2H2; 2.  
DR PROSITE; PS50097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 2.  
KW DNA-binding; Metal-binding; Nuclear protein; Repeat; Transcription;  
KW Transcription regulation; Zinc; Zinc finger.  
FT CHAIN 1 512  
FT FTID=PRO\_0000047721.  
FT FT  
FT DOMAIN 24 92  
FT ZN\_FING 341 363  
FT ZN\_FING 369 392  
FT COMPBTAS 132 153 Poly-Ala.  
SQ SEQUENCE 512 AA; 55562 MW; BB970D200B60CC9F CRC64;  
Query Match 87.0%; Score 47; DB 1; Length 512;

Best Local Similarity 92.3%; Pred. No. 1.6e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAAA 13  
DB 136 AAAAAAAAAAAAAA 148  
RESULT 27  
Q7PPSO ANOQA PRELIMINARY; PRT; 521 AA.  
ID Q7PPSO ANOQA PRELIMINARY; PRT; 521 AA.  
AC Q7PPSO ANOQA PRELIMINARY; PRT; 521 AA.  
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.  
DT 07-DEC-2004, sequence version 2.  
DT 07-FEB-2006, entry version 16.  
DE ENSANGP0000016235.  
GN ORFNames=ENSANGG00000013746;  
OS Anopheles gambiae str. PEST;  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;  
OC Anophelinae; Anopheles.  
OX NCBI TaxID=180454;  
RN [1] NUCLEOTIDE SEQUENCE.  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PEST;  
RG The Anopheles gambiae Sequence Committee;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AA01008933; EAA09944.3; -; Genomic DNA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR InterPro; IPR012677; a\_b\_plait\_nuc\_bd.  
DR InterPro; IPR006532; POLYU\_bd.  
DR InterPro; IPR000504; RNP1\_RNA\_bd.  
DR InterPro; IPR003954; RNP1.  
DR PANTHER; PTHR10432:SF141; PolyU\_half\_pint; 2.  
DR Pfam; PF00076; RRM\_1; 3.  
DR SMART; SM00360; RRM; 3.  
DR SMART; SM00361; RRM\_1; 1.  
DR TIGRPFAMs; TIGR01645; half-pint; 1.  
DR PROSITE; PS50102; RRM; 3.  
SQ SEQUENCE 521 AA; 55760 MW; 68A0C7DEA9FE24DD CRC64;  
Query Match 87.0%; Score 47; DB 2; Length 521;  
Best Local Similarity 92.3%; Pred. No. 1.6e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAAA 13  
DB 339 AAAAAAAAAAAAAA 351  
RESULT 28  
Q4P290 USTWA PRELIMINARY; PRT; 529 AA.  
ID Q4P290 USTWA PRELIMINARY; PRT; 529 AA.  
AC Q4P290;  
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
DT 19-JUL-2005, sequence version 1.  
DT 07-FEB-2006, entry version 4.  
DE Hypothetical protein.  
GN ORFNames=UM05773.1;

```
OS Ustilago maydis 521.
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=237631;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=521;
RA Birren B.W., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachi H.M., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayat T., Blitshteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhalil B., Brunache A., Butler J., Calixte N.,
RA Calvo S.E., Camarata J., Campo K., Chang J., Cheshatsang Y.,
RA Citroen M., Collamore A., Considine T., Cook A., Cooke P., Corum B.,
RA Cuomo C., David R., Dawoe T., Degray S., Dodge S., Dooley K.,
RA Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,
RA Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J.E., Gearin G., Gnerre S.,
RA Gnrke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kanvisselis M., Karlsson E.,
RA Kells C., Kieu A., Kiser P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-Toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.-J., Mabbitt R., MacDonald J., MacLean C., Major J.,
RA Manning J., Marabell R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., McGhee T., Meldrum J., Meneus L.,
RA Mesirov J., Mihalev A., Minova T., Mikkelsen T., Mlenga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., News C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C.B., Nizzari M., Norbu C.,
RA Norbu N., O'Donnell P., Okoawo O., O'Leary S., Omotosho B.,
RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Ratta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schubach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuilinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,
RA Towey S., Tsamila T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V.S., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.S.;
RT "The genome sequence of Ustilago maydis."
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NonDerivs License
CC -----
CC EMBL: AACP01000212; EAK86012.1; -; Genomic_DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003700; P:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam: PF00320; GATA; 1.
DR InterPro: IPR000679; Znf_GATA.
DR SMART: SM00401; Znf_GATA; 1.
DR PROSITE: PS00114; GATA_ZN_FINGER_2; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 529 AA; 53820 MW; D5ABC2C8EA96B06D CRC64;

Query Match      87.0%; Score 47; DB 2; Length 529;
Best Local Similarity 92.3%; Pred. No. 1.6e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAFRAAAAAAAA 13
Db 76 AAVAAAAAAA 88
```



RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
[4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=Berkeley; TISSUE=Embryo;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R.A., Gonzalez M., Guarin H., Krommiller B., Li P.W., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J.M., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K.H., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.E.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
[5]  
RN POSSIBLE FUNCTION, AND TISSUE SPECIFICITY.  
RP PubMed=14597202; DOI=10.1016/j.ydbio.2003.07.011;  
EX Hao I., Green R.B., Dunaevsky O., Lengyel J.A., Rauskolb C.;  
RA "The odd-skipped family of zinc finger genes promotes *Drosophila* leg  
RT segmentation";  
RL Dev. Biol. 263:282-295(2003).  
[6]  
RP TISSUE SPECIFICITY.  
RX MEDLINE=22930851; PubMed=14568103; DOI=10.1016/j.mod.2003.08.001;  
RA Johansen K.A., Green R.B., Iwaki D.D., Hernandez J.B., Lengyel J.A.;  
RT "The Dm-Bowl-Lin relief-of-repression hierarchy controls fore- and  
RL hindgut patterning and morphogenesis";  
RL Mech. Dev. 120:1139-1151(2003).  
CC -!- FUNCTION: Pair-rule protein that determines both the size and  
CC polarity of even-numbered as well as odd-numbered parasegments  
CC during embryogenesis. DNA-binding transcription factor that acts  
CC primarily as a transcriptional repressor but can also function as  
CC a transcriptional activator, depending on the stage of development  
CC and spatial restrictions (by similarity). May function redundantly  
CC with odd and dm in leg joint formation during the larval stages,  
CC acting downstream of Notch activation.  
CC -!- SUBCELLULAR LOCATION: Nucleus (Probable).  
CC -!- TISSUE SPECIFICITY: Has two temporally distinct modes of  
CC expression during early embryogenesis; expressed in seven stripes  
CC at the blastoderm stage. Also expressed in a non-periodic domain  
CC at the anterior of the embryo. During gastrulation, the seven  
CC primary stripes are supplemented by seven secondary stripes that  
CC appear in alternate segments. This results in the labelling of  
CC each of the 14 segments in the extended germ band. Expression is  
CC relatively weak at the blastoderm stage, gaining in intensity at  
CC gastrulation. Expressed in the invaginating stomodaeum and  
CC proctodaeum of the embryonic gut. By stage 13, expressed in the  
CC region that will form the proventriculus and in a wide ring at the  
CC most posterior portion of the midgut. Expression continues in the  
CC gut through the remainder of embryogenesis. Expressed in the  
CC proximal Malpighian tubules, brain and pharyngeal muscles during  
CC late embryogenesis. Expressed weakly in a segmentally repeated  
CC pattern in the leg disk at the distal edge of each presumptive leg  
CC segment except in tarsal segments 1 to 4.  
CC -!- SIMILARITY: Contains 5 C2H2-type zinc fingers.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC  
CC -----  
CC EMBL: U62004; AAC47282.1; -; mRNA.  
CC EMBL: AE003579; AAF51087.1; -; Genomic DNA.  
CC EMBL: BT003205; AAO24960.1; -; mRNA.  
CC PIR: S72227; S72227.  
CC HSSP: P07248; 2ADR.  
CC  
CC FlyBase: FBgn0004892; sob.  
CC DR GO:0005634; C:nucleus; ISS.  
CC DR GO:0003677; F:DNA binding; ISS.  
CC DR GO:0016563; F:transcriptional activator activity; ISS.  
CC DR GO:0016564; F:transcriptional repressor activity; ISS.  
CC DR GO:0007350; P:blastoderm segmentation; ISS.

DR GO:0016348; P:leg joint morphogenesis (sensu Endopterygota); IMP.  
DR GO:0000122; P:negative regulation of transcription from R.; ISS.  
DR GO:0007366; P:periodic partitioning by pair rule gene; ISS.  
DR GO:0045944; P:positive regulation of transcription from R.; ISS.  
DR InterPro: IPR007087; Znf.C2H2.  
DR Pfam: PF00096; zf-C2H2; 5.  
DR ProDom: PD000003; Znf.C2H2; 2.  
DR SMART: SM00355; Znf.C2H2; 5.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 5.  
KW Activator; Complete proteome; Developmental protein; DNA-binding;  
KW Metal-binding; Nucleic acid binding; Pair-rule protein; Repressor;  
KW Transcription; Transcription regulation; Zinc; Zinc-finger.  
FT CHAIN 1 578 Protein sister of odd and bowel.  
FT /FTID=PRO\_0000046928.  
FT ZN\_FING 395 417 C2H2-type 1.  
FT ZN\_FING 423 445 C2H2-type 2.  
FT ZN\_FING 451 473 C2H2-type 3.  
FT ZN\_FING 479 501 C2H2-type 4.  
FT ZN\_FING 507 529 C2H2-type 5.  
FT COMBIAS 9 303 Ala-rich.  
FT COMFLICT 182 256 Ser-rich.  
FT CONFLICT 176 176 S -> P (in Ref. 1).  
FT CONFLICT 188 189 SS -> G (in Ref. 1).  
FT CONFLICT 546 546 P -> L (in Ref. 1).  
SQ SEQUENCE 578 AA; 58455 MW; 0F600954CFA7D8D0 CRC64;  
Query Match 87.0%; Score 47; DB 1; Length 578;  
Best Local Similarity 92.3%; Pred. No. 1.8e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAFAAAAA 13  
Db 274 AAAAAA 286  
RESULT 30  
Q9NVR0 HUMAN PRELIMINARY; PRT; 708 AA.  
AC Q9NVR0;  
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2000, sequence version 1.  
DT 07-MAR-2006, entry version 27.  
DE Hypothetical protein FLJ10572 (Kelch-like 11).  
GN Name=KLHL1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]\_TaxID=9606;  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=14702039; DOI=10.1038/ngl1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka K., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamiyama K., Katsura N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musahino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,





RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gofjohori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Knapp D., Brusic V., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grastrom S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai J., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltas L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reid J.J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.,  
 RA "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573 (2002).  
 [5]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gofjohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690 (2001).  
 [6]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630 (2000).  
 [7]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuwa T., Tashiro H., Itoh M.,  
 RA Suni M., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Iehikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,

RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-Format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771 (2000).  
 [8]  
 RN NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki S., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC  
 DR EMBL; AK045449; BAC32375.1; --; mRNA.  
 DR HSSP; Q14839; IMM2.  
 DR SMR; Q8BR71; 464-519.  
 DR Ensembl; ENSMUSG0000018474; Mus musculus.  
 DR MGI; MGI:1344395; Chd3.  
 DR GO; GO:0016585; C:chromatin remodeling complex; IDA.  
 DR GO; GO:0016581; C:NURD complex; IDA.  
 DR GO; GO:0005515; F:protein binding; IPI.  
 DR GO; GO:0006333; P:chromatin assembly or disassembly; IDA.  
 DR InterPro; IPR0012958; CHD N.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR Pfam; PF08073; CHDNT; 1.  
 DR Pfam; PF00385; Chromo; 1.  
 DR Pfam; PF00628; PHD; 2.  
 DR SMART; SM00298; CHROMO; 2.  
 DR SMART; SM00249; PHD; 2.  
 DR PROSITE; PS00598; CHROMO\_1; UNKNOWN\_1.  
 DR PROSITE; PS0013; CHROMO\_2; 2.  
 DR PROSITE; PS0016; ZF\_PHD\_2; 2.  
 KW Nuclear protein.  
 FT NON\_TER 708 708  
 SQ SEQUENCE 708 AA; 80411 MW; A9CE144CE3310CBB CRC64;  
 QY  
 QY Query Match 87.0%; Score 47; DB 2; Length 708;  
 QY Best Local Similarity 92.3%; Pred. No. 26+02;  
 DB Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 DB 1 RAFAAAAAAAAAA 13  
 DB 218 AAVAAAAAAAAA 230  
 RESULT 32  
 Q8CE33 MOUSE  
 ID Q8CE33\_MOUSE PRELIMINARY; PRT; 709 AA.  
 AC Q8CE33;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 20.  
 DE 10 days neonate skin cDNA, RIKEN full-length enriched library,  
 DE clone:4732491L07 product:hypothetical BTB/POZ domain|Kleish repeat  
 DE 1 antifreeze|Alanine-rich region containing protein, full insert  
 DE sequence.  
 GN Name=Khlh11;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;

RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44 (1999).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Futsaki S., Gariboldi M.,  
 Davis M.J., Wilming L.G., Aizawa K., Aguiar Y., Allen J.E.,  
 Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 Chiu K.P., Choudhary V., Christoffels A., Clutier D.R.,  
 Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 Di Bernardo C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 Kurochkin I.V., Laveau L.F., Lazarevic D., Lipovich L., Liu J.,  
 Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 Mottagui-Tabar S., Mulder N., Nakano N., Nakaguchi H., Ng P.,  
 Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 Petrosky N., Piazza S., Reid J., Reid J.F., Ring B.Z., Ringwald M.,  
 Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
 Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 Grimmond S.M., Teasdale R.D., Liu E.F., Brusic V., Quackenbush J.,  
 Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki K., Tomaru Y.,  
 Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563 (2005).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RA Riken Genome Exploration Research Group, and Genome Science Group  
 (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566 (2005).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 Blake J.A., Bradt T., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
 Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Brownstein M.J., Bult C., Fletcher C., Fujita M., de Bonaldo M.F.,  
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,  
 Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690 (2001).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630 (2000).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=20330913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 Konno H., Akiyama J., Nishi K., Kitsonai T., Tashiro H., Itoh M.,  
 Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 Fujisawa S., Inoue K., Togawa Y., Izawa M., Ohara E., Watabiki M.,  
 Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsushizaki Y.,  
 Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771 (2000).  
 RN RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
 Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Onosato N., Okazaki Y.,  
RA Saito R., Sato H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Takagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tonaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AK029101; BAC26297.1; -; mRNA  
DR Ensembl: ENSMUSG0000048732; Mus musculus.  
DR MGI: 2388648; Khlh1.  
DR GO: GO:0005515; F:protein binding; IEA.  
DR InterPro: IPR011705; BACK.  
DR InterPro: IPR000210; BTB.  
DR InterPro: IPR013069; BTB\_POZ.  
DR InterPro: IPR006652; Kelch\_rep.  
DR Pfam: PF07707; BACK; 1.  
DR Pfam: PF00651; BTB; 1.  
DR Pfam: PF01344; Kelch 1; 3.  
DR SMART: SM00225; BTB; 1.  
DR PROSITE: PS00097; BTB; 1.  
DR Hypothetical protein.  
KW SEQUENCE 709 AA; 80429 MW; BA3DB8D4CA6FEB7A CRC64;  
SQ  
Query Match 87.0%; Score 47; DB 2; Length 709;  
Best Local Similarity 92.3%; Pred No. 2e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 AAFRAAAAAAAAAA 13  
Dy 3 AAFAAAAAAAAAAA 15  
  
RESULT 33  
PYGO DROME  
ID PYGO DROME STANDARD; PRT; 815 AA.  
AC Q9V9W8;  
DT 19-SEP-2002, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAY-2000, sequence version 1.  
DT 21-FEB-2006, entry version 39.  
DE protein pygopus (Gummy legs protein).  
GN Name=PYGO; Synonyms=gam; ORFNames=CG11518;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21952490; PubMed=11955446; DOI=10.1016/S0092-8674(02)00679-7;  
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
RA Murne M., Zuellig S., Basler K.;  
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
RT pygopus to the nuclear beta-catenin-TCF complex.";  
RL Cell 109:47-60(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22010046; PubMed=12015286;  
RA Parker D.S., Jemison J., Cadigan K.M.;  
RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling  
RT in Drosophila.";  
RL Development 129:2565-2576(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Dustin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svicskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
"The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter A.J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=Berkley; TISSUE=Embryo;  
RX MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarni H., Kronmiller B., Pacle J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celisner S.E.;  
RT "A Drosophila full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
CC -!- FUNCTION: Involved in signal transduction through the Wnt pathway.  
CC -!- SUBUNIT: Binds to BCL9 via the PHD-type zinc finger motif, and  
CC thereby becomes part of the nuclear ARM/PAN complex.  
CC -!- SUBCELLULAR LOCATION: Nucleus.  
CC -!- TISSUE SPECIFICITY: Ubiquitous throughout embryogenesis and larval  
CC development.  
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
CC throughout development.  
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AF457206; AAL91369.1; -; mRNA.  
DR EMBL: AY075095; AAL79357.1; -; mRNA.  
DR EMBL: AE003778; AAF57161.1; -; Genomic\_DNA.

DR EMBL; AY058500; AAL13729.1; -; mRNA.  
 DR Ensembl; CG11518; Drosophila melanogaster.  
 DR FlyBase; FBgn0043900; pygo.  
 DR BIOCyc; DMEL-XXX-02:DMEL-XXX-02-014325-MONOMER; -.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:000528; P:transcriptional regulator activity; IPI.  
 DR GO; GO:0035214; P:eye-antennal disc development; IMP.  
 DR GO; GO:0030177; P:positive regulation of Wnt receptor signaling; IPI.  
 DR GO; GO:0007367; P:segment polarity determination; IMP.  
 DR GO; GO:0016055; P:Wnt receptor signaling pathway; IMP.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR Pfam; PF00628; PHD; 1.  
 DR SMART; SM00249; PHD; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE; PS00016; ZF\_PHD\_2; 1.  
 KW Complete proteome; Developmental protein; Metal-binding;  
 KW Nuclear protein; Segmentation polarity protein; Wnt signaling pathway;  
 KW Zinc; Zinc-finger. 815  
 FT CHAIN 1  
 FT ZN FING 747 805  
 FT MOTIF 39 45  
 FT COMPBIAS 48 65  
 FT COMPBIAS 123 749  
 FT CONFLICT 393 393  
 SQ SEQUENCE 815 AA; 80493 MW; 369FDS45D34BC136 CRC64;  
 Query Match 87.0%; Score 47; DB 1; Length 815;  
 Best Local Similarity 92.3%; Pred. No. 2.3e+02;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAFAAAAA 13  
 Db 53 AA VAAAAA 65  
 RESULT 34  
 Q3UG42 MOUSE  
 ID Q3UG42 MOUSE PRELIMINARY; PRT; 828 AA.  
 AC Q3UG42  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 11-OCT-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE B16 F10Y cells cDNA, RIKEN full-length enriched library,  
 DE clone:G370031D15 product:SH3 multiple domains 2, full insert sequence.  
 GN Name=Sh3md2;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning."  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama K., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Allen J.E.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,

RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Huminicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelle J., Kitamura H.,  
 RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Motagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
 RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome."  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RT (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome."  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RN [5]

```
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glasi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [7]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J;
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK148137; BAE28367.1; -; mRNA.
DR MGI; MGI:1913066; Sh3md2.
DR GO; GO:0030027; C:lamellipodium; IDA.
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0046328; F:regulation of JNK cascade; IDA.
DR InterPro; IPR00108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3_1; 3.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRODOM; PD000066; SH3; 3.
DR SMART; SM00184; RING; 1.
DR SMART; SM00326; SH3; 3.
```

```
DR PROSITE; PS50002; SH3; 3.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 828 AA; 86192 MW; 72FE525B907FE0FB CRC64;
Query Match 87.0%; Score 47; DB 2; Length 828;
Best Local Similarity 92.3%; Pred. No. 2.3e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAFAAAAA 13
DB 419 AAVAAAAA 431
RESULT 35
QY 070254 MOUSE PRELIMINARY; PRT; 892 AA.
AC 070254;
DT 01-AUG-1998, integrated into UniProtKB/TrEMBL.
DT 01-AUG-1998, sequence version 1.
DE 07-FEB-2006, entry version 26.
DE Plenty of SH3s.
GN Name-Sh3md2; Synonyms=POSH;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98151363; PubMed=9482736; DOI=10.1093/emboj/17.5.1395;
RA Tapon N., Nagata K., Lamarche N., Hall A.;
RT "A new rac target POSH is an SH3-containing scaffold protein involved
in the JNK and NF-kappaB signalling pathways.";
RL EMBO J. 17:1395-1404(1998).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF030131; AAC40070.1; -; mRNA.
DR FIR; T09071; T09071.
DR HSP; P19878; 1K4U.
DR MGI; MGI:1913066; Sh3md2.
DR GO; GO:0030027; C:lamellipodium; IDA.
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0046328; F:regulation of JNK cascade; IDA.
DR InterPro; IPR00108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3_1; 4.
DR Pfam; PF00097; zf-C3HC4; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRODOM; PD000066; SH3; 4.
DR SMART; SM00184; RING; 1.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS50002; SH3; 4.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
SQ SEQUENCE 892 AA; 93435 MW; 5A1030857777B3B0 CRC64;
Query Match 87.0%; Score 47; DB 2; Length 892;
Best Local Similarity 92.3%; Pred. No. 2.4e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAFAAAAA 13
DB 419 AAVAAAAA 431
RESULT 36
QY 09VE88_DROME
```

ID Q3VE88 DROME PRELIMINARY; PRT; 897 AA.  
 AC Q3VE88;  
 DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
 DT 01-OCT-2002, sequence version 2.  
 DE 21-FEB-2006, entry version 28.  
 DE CG15803-PA.  
 GN -Name=CG15803; ORFNames=Dmel\_CG15803;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.C., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöckle A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Wu D.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RP [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacle B.J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RP [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kanink J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirkas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 a genomics perspective."  
 RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
 RP [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotating the Drosophila melanogaster euchromatic genome: a  
 systematic review."  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RP [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Berkeley Drosophila Genome Project;  
 RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
 RA Hoskins R., Stapleton M., Pacle B.J., Park S., Svirkas R., Smith E.,  
 RA Yu C., Rubin G.;  
 RT "Drosophila melanogaster release 4 sequence."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
 RP [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RG FlyBase;  
 RL Submitted (JAN-2006) to the EMBL/GenBank/DBSJ databases.  
 CC -!- INTERACTION:  
 CC Q9VBP2:CG5053; NDExp=1; IntAct=EBI-172540, EBI-202525;  
 CC Q9VBP7:ed; NDExp=1; IntAct=EBI-172540, EBI-85823;  
 CC -!- SIMILARITY: Contains 4 PDZ (DHR) domains.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL: AE003721; AAF55539.2; -; Genomic\_DNA.  
 DR HSP: Q64512; 1021.  
 DR InAct: Q9VE88;  
 DR FlyBase: Fgn0038606; CG15803.  
 DR GO: GO:0005515; F:protein binding; IPI.  
 DR InterPro: IPR001478; PDZ.  
 DR Pfam: PF00595; PDZ; 4.  
 DR SMART: SM00228; PDZ; 4.  
 DR PROSITE: PS0106; PDZ; 4.  
 SQ SEQUENCE 897 AA; 94372 MW; 3A74789E08FDIA2F CRC64;  
 Query Match 87.0%; Score 47; DB 2; Length 897;  
 Best Local Similarity 92.3%; Pred. No. 2.5e+02;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAFAAAAAAAAAA 13  
 DB 133 AAAAAAAAAAAAA 145  
 RESULT 37  
 Q69211 MOUSE  
 ID Q69211 MOUSE PRELIMINARY; PRT; 914 AA.  
 AC Q69211;  
 DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.  
 DT 13-SEP-2004, sequence version 1.  
 DT 07-FEB-2006, entry version 12.  
 DE MKIAA1494 protein (Fragment).  
 DE Name=Sh3md2; Synonym=mkIAA1494;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryonic tail;  
 RX PubMed=15368895; DOI=10.1093/dnares/11.3.205;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,



```

RA Saga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RA "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT IV. The complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 11:205-218(2004).
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -!- SIMILARITY: Contains 4 SH3 domains.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AK173185; BAD32463.1; -; mRNA.
DR MGI; MGI:1913066; SH3md2.
DR GO; GO:0030027; C:lamellipodium; IDA.
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0046328; F:regulation of JNK cascade; IDA.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001841; Znf_RING.
DR Pfam; PF00018; SH3_1; 4.
DR Pfam; PF00097; zf_C3HC4; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 4.
DR SMART; SM00184; RING; 1.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS00002; SH3; 4.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Metal-binding; SH3 domain; Zinc; Zinc-finger.
FT NON TER 1
SQ SEQUENCE 914 AA; 95963 MW; 0D3F7A4B4AD9E067 CRC64;

Query Match 87.0%; Score 47; DB 2; Length 914;
Best Local Similarity 92.3%; Pred. No. 2.5e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAAAAAAAA 13
DB 441 AAVAAAAAAAAAAA 453

RESULT 38
Q3UYA4 MOUSE PRELIMINARY; PRT; 935 AA.
AC Q3UYA4;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Adult male medulla oblongata cDNA, RIKEN full-length enriched library,
DE clone:630596E803 product:hypothetical Transforming protein Ski/SAND-
DE like/Putative DNA binding containing protein, full insert sequence.
GN Name=Lbxcor1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,

```

```

RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Atsushi R.N., Bailey T.L.,
RA Banael M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gusicich S., Hatzis M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humiński L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Akashira T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno K., Nakano K., Nimomiya N.,
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RN Science 309:1559-1563(2005).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6J; TISSUE=Medulla oblongata;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Oatono N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai J., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,

```



RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.",  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1101/1038(35055500);  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischnmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baidarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=20499174; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Nishimura N.,  
RA Nishiyori H., Nomura K., Ono M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.,  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NonCommercial License  
CC -----  
CC EMBL: AK134840; BAE22308.1; -; mRNA.  
DR MGI: 2443473; Lbxcor1.  
DR GO: 0005634; C:nucleus; RCA.  
DR GO: 0005634; C:nucleus; IDA.  
DR GO: 0005667; C:transcription factor complex; IDA.  
DR GO: 0005515; P:protein binding; IPI.

DR GO: 0003714; F:transcription corepressor activity; IDA.  
DR GO: 0050875; P:cellular physiological process; RCA.  
DR GO: 0016481; P:negative regulation of transcription; IDA.  
DR InterPro: IPR003380; Transform\_Ski.  
DR Pfam: PF02437; Ski\_Sno; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 935 AA; 96933 MW; A9C2953FA477E367 CRC64;  
  
Query Match 87.0%; Score 47; DB 2; Length 935;  
Best Local Similarity 92.3%; Pred. No. 2.5e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAFAAAAA 13  
Db 484 AAVAAAAA 496  
|||||  
|||  
  
RESULT 39  
QSVI93 DROME PRELIMINARY; PRT; 946 AA.  
ID QSVI93 DROME PRELIMINARY; PRT; 946 AA.  
AC QSVI93;  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 2.  
DT 21-FEB-2006, entry version 31.  
DE CG32466-PA, isoform A.  
DN Name=nr; ORFNames=Dmel\_CG32466;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RN Science 287:2185-2195(2000).  
RN [2]

RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RX FlyBase;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC -----  
CC EMBL: AE003672; AAF54032.3; -; Genomic\_DNA.  
DR FlyBase; FBgn0002263; rn.  
DR GO; GO:0007456; P:eye development (sensu Endopterygota); IMP.  
DR GO; GO:0007480; P:leg morphogenesis (sensu Endopterygota); IMP.  
DR InterPro; IPR007087; Znf C2H2.  
DR Pfam; PF000096; zf-C2H2; 6.  
DR ProDom; PD000003; Znf C2H2; 3.  
DR SMART; SM00355; Znf C2H2; 6.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 6.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 6.  
DR Metal-binding; Nuclear protein; Zinc; Zinc-finger.  
SQ SEQUENCE 946 AA; 100948 MW; 05EA01C8F64F61A3 CRC64;

Query Match 87.0%; Score 47; DB 2; Length 946;  
Best Local Similarity 92.3%; Pred. No. 2.6e+02;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAAAA 13  
Db 275 AAAAAAAAAAAAA 287

RESULT 40

LBXCO MOUSE  
ID LBXCO\_MOUSE STANDARD; PRT; 964 AA.  
AC Q8BX46; Q5W812; Q8C0T2;  
DT 21-JUN-2005, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE Ladybird homeobox corepressor 1 (Transcriptional corepressor Corl1).  
GN Names=Ibxcor1; Synonyms=Corl1;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2), FUNCTION, SUBCELLULAR  
RP LOCATION, TISSUE SPECIFICITY, AND INTERACTION WITH LBX1.  
RP TISSUE=Embryonic brain;  
RX PubMed=15528197; DOI=10.1074/jbc.M411652200;  
RA Mizuhara E., Nakatani T., Minaki Y., Sakamoto Y., Ono Y.;  
RT "Corl1, a novel neuronal lineage-specific transcriptional corepressor  
RT for the homeodomain transcription factor Lbx1.";  
RL J. Biol. Chem. 280:3645-3655(2005).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 1 AND 3).  
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Imbomato A.,  
RA Ambesi-Imbomato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels B., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Flecher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Hummel L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tamaoka J., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu B.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plesky C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
CC -!- FUNCTION: Acts as a transcriptional corepressor of LBX1.  
CC -!- SUBUNIT: Interacts with LBX1.  
CC -!- SUBUNIT: Interacts with LBX1.  
CC Q9QVT4:- (xeno); NbExp=1; IntAct=EBI-604451, EBI-604706;  
CC O88712:Ctbp1; NbExp=1; IntAct=EBI-604451, EBI-604547;  
CC O09106:Hdac1; NbExp=1; IntAct=EBI-604451, EBI-301912;  
CC P52955:Ibx1; NbExp=1; IntAct=EBI-604451, EBI-604594;  
CC Q62440:Tle1; NbExp=1; IntAct=EBI-604451, EBI-604471;

```
CC -!- SUBCELLULAR LOCATION: Nucleus.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q8BX46-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=2;
CC IsoId=Q8BX46-2; Sequence=VSP_014179;
CC Name=3;
CC IsoId=Q8BX46-3; Sequence=VSP_014178, VSP_014180, VSP_014181;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Expressed in brain with higher levels in
CC embryo than adult. Also expressed in adult testis. In embryonic
CC brain, expressed in a subset of postmitotic neurons generated
CC posterior to the midbrain-hindbrain border. In the developing
CC spinal cord, selectively expressed in dorsal horn interneurons.
CC -!- SIMILARITY: Belongs to the SKI family.
CC -----
CC Copyrighted by the Uniprot Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB185113; BAD69568.1; -; mRNA.
DR EMBL; AK029916; BAC26674.1; -; mRNA.
DR EMBL; AK049035; BAC33520.1; -; mRNA.
DR HSSP; P12755; lMK1.
DR DR IntAct; Q8BX46; -.
DR DR MGI; MGI:2443473; lbxcor1.
DR DR GO; GO:0005634; C:nucleus; IDA.
DR DR GO; GO:0005667; C:transcription factor complex; IDA.
DR DR GO; GO:0005515; F:protein binding; IPI.
DR DR GO; GO:0003714; F:transcription corepressor activity; IDA.
DR DR GO; GO:0016481; P:negative regulation of transcription; IDA.
DR DR InterPro; IPR003380; Transform_Ski.
DR DR Pfam; PF02437; Ski_Sno; 1.
KW Alternative splicing; Coiled coil; Nuclear protein; Repressor;
KW Transcription; Transcription regulation.
FT CHAIN 1 964 Ladybird homeobox corepressor 1.
FT FTId=PRO_0000129391.
FT COILED 853 921 Potential_
FT VARSPLIC 1 143 Missing (in isoform 3).
FT FTId=VSP_014178.
FT VARSPLIC 9 36 Missing (in isoform 2).
FT FTId=VSP_014179.
FT VARSPLIC 282 290 LQGGGGGA-->PARGRRRR (in isoform 3).
FT FTId=VSP_014180.
FT VARSPLIC 291 964 Missing (in isoform 3).
FT FTId=VSP_014181.
FT CONFLICT 961 961 Q -> P (in Ref. 1).
FT SEQUENCE 964 AA; 100276 MW; BLECDB443D789E2C CRC64;
Query Match 87.0%; Score 47; DB 1; Length 964;
Best Local Similarity 92.3%; Pred. No. 2.6e+02;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAFAAAAAAAAA 13
Db 512 AAFAAAAAAAAA 524
Search completed: September 9, 2006, 23:00:21
Job time : 120.139 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:37:17 ; Search time 97.4177 Seconds  
(without alignments)  
61.014 Million cell updates/sec

Title: US-10-617-568-5

Perfect score: 54

Sequence: 1 ASMSAASAAASWAA 13

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A\_Geneseq\_8.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*
- 10: Geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	13	ADI29009	ADI29009 Human MHC
2	42	77.8	417	ABB70883	Abb70883 Drosophil
3	42	77.8	970	ABP27418	Abp27418 Streptoco
4	42	77.8	1310	ADU69631	Adu69631 S agalat
5	42	77.8	1310	ADV89548	Adv89548 Streptoco
6	42	77.8	1310	ADV80801	Adv80801 Streptoco
7	42	77.8	1310	ADV82959	Adv82959 Streptoco
8	40	74.1	510	ABB64821	Abb64821 Drosophil
9	39	72.2	276	ADI60138	Adi60138 Secreted
10	39	72.2	311	ADM17198	Adm17198 E grandis
11	39	72.2	395	ADI42807	Adi42807 Plant tra
12	39	72.2	395	ADO03010	Ado03010 Thalecres
13	39	72.2	395	ADO62431	Ado62431 Transcrip
14	39	72.2	793	ABB58866	Abb58866 Drosophil
15	39	72.2	930	ABM89198	Abm89198 Rice abio
16	38	70.4	12	ADI29006	ADI29006 Human MHC
17	38	70.4	328	ABB49912	Abb49912 Listeria
18	38	70.4	508	ABB71345	Abb71345 Drosophil
19	38	70.4	589	ABR40698	AbR40698 Glycine m
20	37	68.5	45	ABM87998	Abm87998 Rice abio
21	37	68.5	170	ABM92683	Abm92683 M. xanthu
22	37	68.5	285	ABB65204	Abb65204 Drosophil
23	37	68.5	375	ABU50372	Abu50372 Protein e

24	37	68.5	424	3	AA558798	Aab558798 Breast an
25	37	68.5	455	8	ADR43053	Adr43053 IPI-like
26	37	68.5	455	10	AEF15580	Aef15580 Isopenten
27	37	68.5	493	6	ABQ71875	Abq71875 Pseudomon
28	37	68.5	533	6	ABU22267	Abu22267 Protein e
29	36	66.7	13	8	ADI29007	ADI29007 Human MHC
30	36	66.7	20	7	ADH47583	Adh47583 Human lun
31	36	66.7	20	7	ADH47584	Adh47584 Human lun
32	36	66.7	20	8	ADJ21502	Adj21502 Human lun
33	36	66.7	20	8	ADJ21503	Adj21503 Human lun
34	36	66.7	148	4	ABB70649	Abb70649 Drosophil
35	36	66.7	250	8	ADT59595	Adt59595 Plant pol
36	36	66.7	268	7	ABO83352	AbO83352 Pseudomon
37	36	66.7	327	8	ADR98917	Adr98917 Lung spec
38	36	66.7	327	4	ABB66932	Abb66932 Drosophil
39	36	66.7	327	4	AAB94055	Aab94055 Human pro
40	36	66.7	327	4	AAM93990	Aam93990 Human sto
41	36	66.7	329	5	ABB50003	Abb50003 Listeria
42	36	66.7	329	6	ABU32725	Abu32725 Protein e
43	36	66.7	344	6	AAE36904	Aae36904 Spider (E
44	36	66.7	367	5	ABB10101	Abb10101 Human hom
45	36	66.7	367	5	ABG70892	Abg70892 Human NKX
46	36	66.7	367	8	ADO09851	Ado09851 Human NK-
47	36	66.7	402	6	ADA48438	Ada48438 Rice prot
48	36	66.7	463	4	ABB70668	Abb70668 Drosophil
49	36	66.7	474	5	AU85550	Au85550 Clone #24
50	36	66.7	474	6	ABU56582	Abu56582 Lung canc
51	36	66.7	474	6	ABU69522	Abu69522 Human lun
52	36	66.7	474	6	ABU66425	Abu66425 Lung canc
53	36	66.7	474	7	ADB75557	Adb75557 Prostate
54	36	66.7	474	7	ADH47331	Adh47331 Human lun
55	36	66.7	474	8	ABO58403	Abo58403 Human gen
56	36	66.7	474	8	ADJ21250	Adj21250 Human lun
57	36	66.7	474	8	ABM80791	Abm80791 Tumour-as
58	36	66.7	474	8	ADX91707	Adx91707 Plant ful
59	36	66.7	474	9	ADW47985	Adw47985 Human sex
60	36	66.7	474	9	ADX05999	Adx05999 Cyclin-de
61	36	66.7	474	9	AEA04475	Aea04475 Human pro
62	36	66.7	482	9	ADO66985	Ado66985 Cyclin-de
63	36	66.7	482	9	ADY70507	Ady70507 Human bet
64	36	66.7	485	6	ABR82252	AbR82252 Human nuc
65	36	66.7	494	4	ABG02308	Abg02308 Novel hum
66	36	66.7	509	2	AAR85311	Aar85311 Cyathathio
67	36	66.7	509	6	ADB23053	AdB23053 Corn cyst
68	36	66.7	537	4	ABB60713	Abb60713 Drosophil
69	36	66.7	568	9	AE827353	Aeb27353 Pinus rad
70	36	66.7	646	8	ADY09213	Ady09213 Plant ful
71	36	66.7	686	1	AAF93711	Aap93711 Sequence
72	36	66.7	734	6	AAE36888	Aae36888 Eugrus c
73	36	66.7	742	4	ABB59853	Abb59853 Drosophil
74	36	66.7	797	4	ABB60378	Abb60378 Drosophil
75	36	66.7	815	4	ABE70499	Abb70499 Drosophil
76	36	66.7	815	6	ABF54340	Abp54340 Drosophil
77	36	66.7	1121	6	ABE63178	Abb63178 Drosophil
78	36	66.7	2016	6	AAB36891	Aae36891 Plectreur
79	36	66.7	2023	4	ABB63487	Abb63487 Drosophil
80	36	66.7	5069	2	AAW52846	Aaw52846 A. medite
81	35	64.8	35	2	AAI16038	Aay16038 Peptide e
82	35	64.8	92	3	AAI44712	Aay44712 Shorthorn
83	35	64.8	130	8	ADX71924	Adx71924 Plant ful
84	35	64.8	135	4	AAU57985	Aau57985 Propionib
85	35	64.8	135	6	ABM54504	Abm54504 Propionib
86	35	64.8	151	8	ADY13585	Ady13585 Plant ful
87	35	64.8	152	8	ADY06421	Ady06421 Plant ful
88	35	64.8	173	4	ABM62452	Abb62452 Drosophil
89	35	64.8	182	7	ABM86593	Abm86593 Rice abio
90	35	64.8	196	8	ADY79232	Ady79232 Plant ful
91	35	64.8	213	7	ABO64613	AbO64613 Klebsiell
92	35	64.8	237	4	ABE68184	Abb68184 Drosophil
93	35	64.8	249	7	ABO70467	AbO70467 Pseudomon
94	35	64.8	264	4	ABE65493	Abb65493 Drosophil
95	35	64.8	272	8	ADX90400	Adx90400 Plant ful
96	35	64.8	291	8	ADI42827	Adi42827 Plant tra

97 35 64.8 291 8 ADO03023 Ado03023 Thalecres  
 98 35 64.8 292 2 AAY37630 Aay37630 Chlamydia  
 99 35 64.8 314 10 AEF77642 Aef77642 Rat 3-hyd  
 100 35 64.8 314 10 AEF77640 Aef77640 Mouse 3-h

## ALIGNMENTS

RESULT 1  
 ADI29009  
 ID ADI29009 standard; peptide; 13 AA.  
 XX AC  
 XX ADI29009;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Human MHC class II compound spacerholder molecule SEQ ID NO:5.  
 XX  
 KW MHC class II compound; MHC class II component; MHC class II alpha chain;  
 KW MHC class II beta chain; peptide binding groove; spacerholder molecule;  
 KW effector component; immune response; immune disorder; virucide;  
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
 KW gene therapy; viral infections; bacterial infection; parasitic infection;  
 KW neoplastic disease; autoimmunity; toxicity; human.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO2004007528-A2.  
 XX  
 PD 22-JAN-2004.  
 XX  
 XX 11-JUL-2003; 2003WO-US021767.  
 XX  
 PR 12-JUL-2002; 2002US-0395494P.  
 PR 22-JUL-2002; 2002US-0397893P.  
 XX  
 PA (DAND ) DANA FARBER CANCER INST INC.  
 XX  
 PI Wucherpfenning KW, Seth N;  
 XX  
 DR WPI; 2004-122876/12.  
 XX  
 XX New MHC class II compound, useful for preparing a composition for  
 PT treating immune disorders e.g. viral infections, bacterial infections,  
 PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
 XX  
 PS Claim 11; SEQ ID NO 5; 92pp; English.  
 XX  
 CC The present invention describes an isolated MHC class II compound (I)  
 CC comprising: (a) an MHC class II component comprising at least a portion  
 CC of an MHC class II alpha chain and at least a portion of an MHC class II  
 CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
 CC chain form a peptide binding groove; (b) a spacerholder molecule; and (c)  
 CC an effector component, where the effector component is linked to the MHC  
 CC class II component. Also described: (1) a pharmaceutical composition  
 CC comprising the MHC class II molecule and a carrier; (2) a method of  
 CC producing an MHC class II compound; (3) a method of directly identifying  
 CC an antigen-specific T cell; (4) a method of regulating an immune response  
 CC in a subject; (5) a method of treating an immune disorder in a subject;  
 CC (6) a method of regulating an immune response ex vivo in a subject; and  
 CC (7) a method of treating an immune disorder ex vivo in a subject. (I) has  
 CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive  
 CC activities, and can be used in gene therapy. The MHC class II compound  
 CC (I) can be used for preparing a composition for treating immune  
 CC disorders, e.g., viral infections, bacterial infections, parasitic  
 CC infections, neoplastic disease, autoimmunity or toxicity. The present  
 CC sequence represents a spacerholder molecule peptide, which can be used in  
 CC an MHC class II compound from the present invention.  
 XX  
 SQ Sequence 13 AA;

Query Match 100.0%; Score 54; DB 8; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 0.024;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ASMSAASAASMAA 13  
 |||||  
 DB 1 ASMSAASAASMAA 13  
 |||||  
 RESULT 2  
 ABB70883  
 ID ABB70883 standard; protein; 417 AA.  
 XX AC  
 XX ABB70883;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 39441.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-658860/75.  
 DR N-PSDB; ABL14986.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 39441; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
 CC ABB2702). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 417 AA;  
 Query Match 77.8%; Score 42; DB 4; Length 417;  
 Best Local Similarity 76.9%; Pred. No. 85;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ASMSAASAASMAA 13  
 |||||  
 DB 225 ASTSAASMAA 237  
 |||||  
 RESULT 3  
 ABP27418  
 ID ABP27418 standard; protein; 970 AA.  
 XX  
 AC ABP27418;

XX 02-JUL-2002 (first entry)  
 DT Streptococcus polypeptide SEQ ID NO 4012.  
 DE Streptococcus; GAS; GBS; Group B streptococcus; Streptococcus agalactiae;  
 DE Group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 KW Streptococcus agalactiae.  
 XX WO200234771-A2.  
 XX 02-MAY-2002.  
 PD 29-OCT-2001; 2001WO-GB004789.  
 XX 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;  
 PI Tetelin H;  
 XX WPI; 2002-352536/38.  
 DR N-PSDB; ABN68049.  
 XX New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX Claim 1; Page 3556; 4525pp; English.  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX Sequence 970 AA;  
 SQ Query Match 77.8%; Score 42; DB 5; Length 970;  
 Best Local Similarity 69.2%; Pred. No. 2.1e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ASMSAASASMAA 13  
 DB 881 ASMSASTASMSA 893  
 RESULT 4  
 ADU69631  
 ID ADU69631 standard; protein; 1310 AA.  
 XX AC ADU69631;  
 XX 10-FEB-2005 (first entry)  
 XX

DE S agalactiae hyperimmune serum reactive antigen seqid 326.  
 XX immune stimulation; antigen; bacterial surface display;  
 KW hyperimmune serum reactive antigen; vaccine; bacterial infection;  
 KW antibacterial; infection.  
 XX Streptococcus agalactiae.  
 XX WO2004099242-A2.  
 PN 18-NOV-2004.  
 PD 06-MAY-2004; 2004WO-EP004856.  
 XX 07-MAY-2003; 2003EP-00450112.  
 PR 28-NOV-2003; 2003EP-00450266.  
 XX (INTE-) INTERCELL AG.  
 PA Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;  
 PI WPI; 2004-821662/81.  
 XX DR N-PSDB; ADU69414.  
 XX New nucleic acid molecule encoding a hyperimmune serum reactive antigen,  
 PT useful for the manufacture of a vaccine against Streptococcus agalactiae  
 PT infection.  
 XX Claim 12; SEQ ID NO 326; 221pp; English.  
 XX The invention describes an isolated nucleic acid molecule encoding a  
 CC hyperimmune serum reactive antigen or its fragment. Also described are: a  
 CC vector comprising the nucleic acid molecule; a host cell comprising the  
 CC vector; a hyperimmune serum-reactive antigen comprising a sequence  
 CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467  
 CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens  
 CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;  
 CC a process for producing a Streptococcus agalactiae hyperimmune serum  
 CC reactive antigen; a process for producing a cell that expresses a S.  
 CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical  
 CC composition, especially a vaccine, comprising the hyperimmune serum-  
 CC reactive antigen or nucleic acid molecule; an antibody that binds at  
 CC least to a selective part of the hyperimmune serum-reactive antigen; a  
 CC hybridoma cell line, which produces the antibody; a method for producing  
 CC the antibody; an antagonist that binds to the hyperimmune serum-reactive  
 CC antigen; a method for identifying an antagonist capable of binding to the  
 CC hyperimmune serum-reactive antigen; a method for identifying an  
 CC antagonist capable of reducing or inhibiting the interaction activity of  
 CC a hyperimmune serum-reactive antigen to its interaction partner; a  
 CC process for in vitro diagnosing a disease related to expression of the  
 CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis  
 CC of a bacterial infection, especially a S. agalactiae infection. The  
 CC hyperimmune serum reactive antigen is useful for isolating, purifying  
 CC and/or identifying an interaction partner of the hyperimmune serum  
 CC reactive antigen. The hyperimmune serum reactive antigen is useful for  
 CC generating a peptide binding to the hyperimmune serum reactive antigen,  
 CC where the peptide comprises anticalines, or for the manufacture of a  
 CC functional nucleic acid comprising aptamers or Spiegelmers. The nucleic  
 CC acid molecule is useful for the manufacture of a functional ribonucleic  
 CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic  
 CC acid molecule, hyperimmune serum-reactive antigen or antibody is useful  
 CC for the manufacture of a vaccine against S. agalactiae infection. This is  
 CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum  
 CC reactive antigen.  
 XX Sequence 1310 AA;  
 SQ Query Match 77.8%; Score 42; DB 8; Length 1310;  
 Best Local Similarity 69.2%; Pred. No. 2.9e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 ASMSAASASMAA 13  
 |||||:::|||||

```

Db      781 ASMSASTSASMSA 793

RESULT 5
ADV89548
ID   ADV89548 standard; protein; 1310 AA.
XX
AC   ADV89548;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus agalactiae protein sequence, SEQ ID 1942.
XX
KW   Antibacterial; Vaccine; bacterial infection.
XX
OS   Streptococcus agalactiae.
XX
PN   WO200292818-A2.
XX
PD   21-NOV-2002.
XX
PF   26-APR-2002; 2002WO-IB003059.
XX
PR   26-APR-2001; 2001FR-00005642.
XX
PA   (INSP ) INST PASTEUR.
(CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;
PI   Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX
DR   WPI; 2004-101891/11.
XX
PT   Genomic nucleotide sequences encoding polypeptides of Streptococcus
PT   agalactiae for the development of vaccines, diagnostic tools, DNA chips
PT   and identification of therapeutic targets.
XX
PS   Claim 6; SEQ ID NO 1942; 439pp; French.
XX
CC   The present invention relates to novel Streptococcus agalactiae
CC   nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV84476) and
CC   novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
CC   nucleotide sequences encode polypeptides of S. agalactiae involved in the
CC   synthesis of amino acids, cell membranes, intermediate (central)
CC   metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
CC   nucleotide metabolism including purines, pyrimidines and/or nucleosides,
CC   regulatory functions, replication, transcription, translation, protein
CC   transport, adaptation to atypical conditions, sensitivity to medicines
CC   and/or analogues, functions related to transposons, biosynthesis of
CC   cofactors, prosthetic groups and transporters, cell membrane proteins and
CC   cellular machinery. (I) are useful for the detection and/or amplification
CC   of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
CC   useful for treatment of a bacterial S. agalactiae infection. The complete
CC   genome of Streptococcus agalactiae is given in ADV81204. Note: The
CC   present patent is an equivalent for the basic patent FR2824074A1, which
CC   contains only 2344 sequences.
XX
SQ   Sequence 1310 AA;
      Query Match      77.8%; Score 42; DB 8; Length 1310;
      Best Local Similarity 69.2%; Pred. No. 2.9e+02;
      Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy      1 ASMSAASASMAA 13
      |||||:::|
Db      781 ASMSASTSASMSA 793

RESULT 7
ADV82959
ID   ADV82959 standard; protein; 1310 AA.
XX
AC   ADV82959;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Streptococcus agalactiae protein, SEQ ID 4100.
XX
KW   Antibacterial; vaccine; bacterial infection.
XX
OS   Streptococcus agalactiae.

```

XX WO200292818-A2.  
 XX PD 21-NOV-2002.  
 XX PF 26-APR-2002; 2002WO-IB003059.  
 XX PR 26-APR-2001; 2001FR-00005642.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX DR WPI; 2004-101891/11.  
 XX CC Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 XX PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 XX PT and identification of therapeutic targets.  
 XX PS Claim 6; SEQ ID NO 4100; 439pp; French.  
 XX CC The present invention relates to novel Streptococcus agalactiae  
 XX CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 XX CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 XX CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 XX CC synthesis of amino acids, cell membranes, intermediate (central)  
 XX CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 XX CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 XX CC regulatory functions, replication, transcription, translation, protein  
 XX CC transport, adaptation to atypical conditions, sensitivity to medicines  
 XX CC and/or analogues, functions related to transposons, biosynthesis of  
 XX CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 XX CC cellular machinery. (I) are useful for the detection and/or amplification  
 XX CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 XX CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 XX CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 XX CC present patent is an equivalent for the basic patent FR282407A1, which  
 XX CC contains only 2344 sequences.  
 XX SQ Sequence 1310 AA;

Query Match 77.8%; Score 42; DB 8; Length 1310;  
 Best Local Similarity 69.2%; Pred. No. 2.9e+02;  
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13  
 |||||:|:|:|:  
 Db 781 ASMSASTSASMSA 793

RESULT 8  
 ABB64821  
 ID ABB64821 standard; protein; 510 AA.  
 XX AC ABB64821;  
 XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster polypeptide SEQ ID NO 21255.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX KW pharmaceutical.  
 XX OS Drosophila melanogaster.  
 XX PN WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX DR WPI; 2001-856860/75.  
 XX DR N-PSDB; ABL08924.  
 XX CC New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 XX PT interactions.  
 XX PS Disclosure; SEQ ID NO 21255; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX CC useful in developmental biology and in elucidating cell signalling and  
 XX CC cell-cell interactions in higher eukaryotes for the development of  
 XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 XX CC sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-  
 XX CC ABB72072). The sequence data for this patent did not form part of the  
 XX CC printed specification, but was obtained in electronic format directly  
 XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 510 AA;

Query Match 74.1%; Score 40; DB 4; Length 510;  
 Best Local Similarity 69.2%; Pred. No. 2.2e+02;  
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ASMSAASASMAA 13  
 |||||:|:|:|:  
 Db 101 ASMSASMSASMSA 113

RESULT 9  
 ADI60138  
 ID ADI60138 standard; protein; 276 AA.  
 XX AC ADI60138;  
 XX DT 15-APR-2004 (first entry)  
 XX DE Secreted polypeptide #22.  
 XX KW osteopathic; vulnery; cytostatic; gene therapy; diagnosis; forensics;  
 KW gene mapping; mutation identification; biodiversity; chromosome marker;  
 KW immune response; myeloid cell disorder; lymphoid cell disorder;  
 KW bone cartilage; tendon; ligament; nerve tissue growth; wound healing;  
 KW burns; incision; ulcer; cancer.  
 XX OS Homo sapiens.  
 XX PN WO2003025142-A2.  
 XX PD 27-MAR-2003.  
 XX PF 18-SEP-2002; 2002WO-US029636.  
 XX PR 18-SEP-2001; 2001US-0323349P.  
 XX PR 16-SEP-2002; 2002US-00323349.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
 XX PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT;  
 XX DR WPI; 2003-354601/33.  
 XX DR N-PSDB; ADI60483.



PT New polynucleotides and secreted proteins, useful for treating myeloid or  
 PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
 PT tissue growth or regeneration, in wound healing, and in tissue repair and  
 PT replacement.

PS Claim 20; SEQ ID NO 173; 243pp; English.

XX The invention relates to novel isolated polynucleotides or a sequence  
 CC encoding a polypeptide with biological activity, where the polynucleotide  
 CC hybridizes to the polynucleotide under stringent hybridization conditions  
 CC or has greater than 95% sequence identity with the polynucleotide. The  
 CC polynucleotides and polypeptides are useful in diagnostics, forensics,  
 CC gene mapping, identification of mutations responsible for genetic  
 CC disorders and other traits, to assess biodiversity, as nutritional  
 CC sources or supplements. The polynucleotides may also be used as molecular  
 CC weight markers, chromosome markers or map related gene positions, or as  
 CC an antigen to raise anti-DNA antibodies or elicit immune response. The  
 CC polypeptides are useful for raising antibodies, as markers for tissues in  
 CC which the corresponding polypeptide is expressed, for re-engineering  
 CC damaged or diseased tissues, for treating myeloid or lymphoid cell  
 CC disorders, in bone cartilage, tendon, ligament and/or nerve tissue growth  
 CC or regeneration, in wound healing, in tissue repair and replacement, in  
 CC healing of burns, incisions and ulcers, and in treating cancer. This  
 CC sequence corresponds to a protein sequence of the invention.

XX Sequence 276 AA;

Query Match 72.2%; Score 39; DB 7; Length 276;

Best Local Similarity 61.5%; Pred. No. 1.7e+02; Indels 0; Gaps 0;

Matches 8; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13

Db 5 AALSAAAAALAA 17

RESULT 10

ID ADW17198 standard; protein; 311 AA.

XX ADW17198;

XX 24-MAR-2005 (first entry)

XX Egrandis transcription factor protein C2C2 (co-like) family Seq 948.

XX plant; transcription; gene regulation; gene expression; transgenic plant;  
 KW drought resistance; disease resistance; salt tolerance; cold tolerance;  
 KW freezing tolerance; flowering; flavor enhancer; flower color.

XX Eucalyptus grandis.

OS WO2005001050-A2.

XX 06-JAN-2005.

XX 07-JUN-2004; 2004WO-US017965.

XX 06-JUN-2003; 2003US-0476189P.

XX (ARBO-) ARBORGEN LLC.

XX Bloksberg LN, Bryant C, Connett MB, Emerson SJ, Frost MJ;

PI Forster RLS, Grigor M, Higgins C, Lasham A, Lund ST, Magusin A;

PI Phillips J, Puthigae S, Veerakoe S, Westwood C, Gause K, Wood M;

XX WPI; 2005-075542/08.

DR N-PSDB; ADW16412.

XX New polynucleotides isolated from plants encoding transcription factors,  
 PT and polypeptides encoded by such polynucleotides, useful for regulating  
 PT gene transcription and gene expression.

PS Claim 31; SEQ ID NO 948; 1265pp; English.

XX This invention relates to novel isolated plant nucleic acid molecules, or  
 CC variants thereof, that encode transcription factors. Specifically, it  
 CC refers to transcription factor proteins that are capable of binding to  
 CC DNA in order to regulate gene transcription and gene expression in a  
 CC plants, in particular Eucalyptus grandis and Pinus radiata. The present  
 CC invention describes DNA constructs containing DNA encoding a  
 CC transcription factor that regulates the promoter, which is operably  
 CC linked to the desired nucleic acid to be expressed. It further provides  
 CC transgenic plants expressing a transcription factor that confers a trait  
 CC to the plant such as increased drought, salt or disease tolerance, height  
 CC change, enhanced cold/ frost tolerance, enhanced color, health and  
 CC nutritional characteristics, as well as improved taste, starch  
 CC composition, flower longevity and germination, amongst others.  
 CC Accordingly, such plants that are successfully transfected with a DNA  
 CC construct can be characterized by a difference in flower color, petal or  
 CC leaf shape and size, aroma or plant height. This polypeptide is a plant  
 CC transcription factor protein sequence of the invention.

XX Sequence 311 AA;

Query Match 72.2%; Score 39; DB 9; Length 311;

Best Local Similarity 76.9%; Pred. No. 1.9e+02;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASAAASMAA 13

Db 155 AAASAAASAAASAA 167

RESULT 11

AD142807

ID AD142807 standard; protein; 395 AA.

XX AD142807;

XX 22-APR-2004 (first entry)

XX Plant transcription factor #459.

XX transgenic; plant; enhanced tolerance to abiotic stress;

KW glycosylase tolerance; hormone sensitivity; disease resistance;

KW sugar sensing; flowering; flower structure; stem bifurcation;

KW branching pattern; apical dominance; trichome; stem morphology;

KW root growth; root hair; seed development; cell proliferation;

KW cell differentiation; premature senescence; necrosis; plant size;

KW leaf morphology; seed morphology; seed biochemistry; root anthocyanin;

KW plant anthocyanin; light response; shade avoidance; bioinformatic;

KW transcription factor; gene; ds.

XX Oryza sativa.

XX US2004019927-A1.

XX 29-JAN-2004.

XX 25-FEB-2003; 2003US-00374780.

XX 18-APR-2001; 2001US-00837944.

XX (SHER/) SHERMAN B K.

PA (RIEC/) RIECHMANN J L.

PA (JIAN/) JIANG C.

PA (HEAR/) HEARD J E.

PA (HAAR/) HAAKE V.

PA (CREE/) CREELMAN R A.

PA (RATC/) RATCLIFFE O.

PA (ADAM/) ADAM L J.

PA (REUB/) REUBER T L.

PA (KEDD/) KEDDIE J.

PA (BROU/) BROUN P E.

PA (PILG/) PILGRIM M L.

```

PA (DUBE//) DUBELL A N.
PA (PINE//) PINEDA O.
PA (YUGG//) YU G.
XX
XX Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V;
PI Creelman RA, Ratcliffe O, Adam LJ, Reuber TL, Keddle J, Broun PE;
PI Pilgrim ML, Dubell AN, Ratcliffe O, Kumimoto R;
XX
XX WPI; 2004-132245/13.
XX
XX New transgenic plant comprising a recombinant polynucleotide of any one
PT of more than 500 nucleotide sequences, useful in bioinformatic search
PT methods.
XX
XX Claim 1; SEQ ID NO 1270; 435pp; English.
XX
XX The invention describes a transgenic plant comprising a recombinant
CC polynucleotide of any one of more than 500 nucleotide sequences fully
CC defined in the specification or its complement. The method of the
CC invention can be used to produced a plant having altered traits such as:
CC enhanced tolerance to abiotic stress; glyphosphate tolerance; hormone
CC sensitivity; disease resistance; sugar sensing; early or late flowering;
CC altered flower structure, change in stem bifurcations, altered branching
CC pattern, reduced apical dominance, reduced trichome density; lack of
CC trichomes; reduced ectopic trichome development; altered trichome
CC development; increase in trichome number; altered stem morphology;
CC increased root growth; increased root hairs; altered seed development;
CC altered cell proliferation or cell differentiation; rapid development;
CC premature senescence; increased necrosis; increase in seedling or plant
CC size; decreased plant size; leaf morphology; seed morphology; seed
CC biochemistry; increase in root anthocyanins; increase in plant
CC anthocyanins, or alteration in light response or shade avoidance. The
CC transgenic plant, polynucleotides and polypeptides are useful in
CC bioinformatic search methods. This is the amino acid sequence of a plant
CC transcription factor, and an orthologue of Arabidopsis thaliana
CC transcription factors isolated in the invention, that can be used in the
CC creation of a transgenic plant with altered traits.
XX
XX Sequence 395 AA;
SQ
Query Match 72.2%; Score 39; DB 8; Length 395;
Best Local Similarity 69.2%; Pred. No. 2.5e+02;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ASMSAASASMAA 13
Db 264 AMSGASAAAAA 276

RESULT 12
ID ADO03010 standard; protein; 395 AA.
XX
XX ADO03010;
AC
XX
XX 01-JUL-2004 (first entry)
DT
XX
XX Thalecress transcription factor, Rice orthologue #82.
DE
XX
XX Rice; transcription factor; plant; transgenic; abiotic stress;
KW cold tolerance; heat tolerance; drought; osmotic stress;
KW phosphate limitation; potassium limitation; nitrogen limitation;
KW hormone sensitivity; disease resistance; sugar sensing; seed germination;
KW flowering; inflorescence architectural change;
KW meristem cell differentiation; phyllotaxy; apical dominance;
KW trichome development; seed development; premature senescence;
KW delayed senescence; lethality; necrosis; plant size; leaf morphology;
KW seed morphology; secondary metabolism; light response; shade avoidance.
XX
XX Oryza sativa.
OS
XX
XX US2004045049-A1.
PN
XX

```

---

```

PD (DUBE//) DUBELL A N.
XX (PINE//) PINEDA O.
XX (YUGG//) YU G.
XX
XX Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V;
PR 21-JAN-2000; 2000US-00489376.
PR 17-FEB-2000; 2000US-00506720.
PR 22-MAR-2000; 2000US-00532591.
PR 22-MAR-2000; 2000US-00533029.
PR 22-MAR-2000; 2000US-00533030.
PR 22-MAR-2000; 2000US-00533392.
PR 22-MAR-2000; 2000US-00533648.
PR 06-APR-2000; 2000WO-US009448.
PR 16-NOV-2000; 2000US-00713994.
PR 27-MAR-2001; 2001US-00819142.
PR 17-APR-2001; 2001US-00837444.
PR 30-JAN-2002; 2002US-00958131.
PR 14-JUN-2002; 2002US-00171468.
PR 09-AUG-2002; 2002US-00225066.
PR 09-AUG-2002; 2002US-00225067.
PR 09-AUG-2002; 2002US-00225068.
PR 17-DEC-2002; 2002US-0434166P.
PR 25-FEB-2003; 2003US-00374780.
XX
XX (ZHAN//) ZHANG J.
PA (FROM//) FROMM M E.
PA (HEAR//) HEARD J E.
PA (RIEC//) RIECHMANN J L.
PA (ADAM//) ADAM L J.
PA (BROU//) BROUN P E.
PA (PINE//) PINEDA O.
PA (REUB//) REUBER T L.
PA (KEDD//) KEDDIE J S.
PA (YUGG//) YU G.
PA (JIANG//) JIANG C.
PA (SAMA//) SAWAHA R S.
PA (PILG//) PILGRIM M L.
PA (CREE//) CREELMAN R A.
PA (DUBE//) DUBELL A N.
PA (RATC//) RATCLIFFE O.
PA (KUMI//) KUMIMOTO R.
PA (SHER//) SHERMAN B K.
XX
XX Zhang J, Fromm MB, Heard JE, Riechmann JL, Adam LJ, Broun PE;
PI Pineda O, Reuber TL, Keddle JS, Yu G, Jiang C, Samaha RS;
PI Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;
PI Sherman BK;
XX
XX WPI; 2004-225755/21.
XX
XX New transgenic plant, useful in developing phenotypes with altered or
PT improved characteristics or traits.
XX
XX Claim 1; SEQ ID NO 1424; 213pp; English.
XX
XX The invention relates to a transgenic plant comprises a recombinant
CC polynucleotide having a polynucleotide sequence or its complementary
CC sequence comprising a sequence encoding a polypeptide, that initiates
CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,
CC Rice, Rape or Corn, comprising any of the sequences appearing as ADO01588
CC -ADO03527 or ADO03530-ADO03559. Also included are using a transgenic
CC plant to grow a progeny plant, an expression cassette (comprising a
CC constitutive, inducible or tissue-specific promoter and a recombinant
CC polynucleotide described above), a host cell comprising the expression
CC cassette, producing a modified plant having a modified trait, identifying
CC a factor that is modulated by or interacts with a polypeptide encoded by
CC the polynucleotide sequence and identifying at least one downstream
CC polynucleotide sequence that is subject to a regulatory effect of any of
CC the polypeptides encoded by the polynucleotide described above. The
CC transgenic plant is useful for producing a plant that has an altered
CC trait e.g. an enhanced tolerance to abiotic stress (increased tolerance
CC to chilling, germination in cold conditions, freezing tolerance, tolerance
CC to heat, tolerance to drought, tolerance to osmotic stress, tolerance to

```





PT treating immune disorders e.g. viral infections, bacterial infections,  
PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
XX  
XX  
PS Claim 11; SEQ ID NO 2; 92pp; English.  
XX  
XX The present invention describes an isolated MHC class II compound (I)  
CC comprising: (a) an MHC class II component comprising at least a portion  
CC of an MHC class II alpha chain and at least a portion of an MHC class II  
CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
CC chain form a peptide binding groove; (b) a spacerholder molecule; and (c)  
CC an effector component, where the effector component is linked to the MHC  
CC class II component. Also described: (1) a pharmaceutical composition  
CC comprising the MHC class II molecule and a carrier; (2) a method of  
CC producing an MHC class II compound; (3) a method of directly identifying  
CC an antigen-specific T cell; (4) a method of regulating an immune response  
CC in a subject; (5) a method of treating an immune disorder in a subject;  
CC (6) a method of regulating an immune response ex vivo in a subject; and  
CC (7) a method of treating an immune disorder ex vivo in a subject. (I) has  
CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive  
CC activities, and can be used in gene therapy. The MHC class II compound  
CC (I) can be used for preparing a composition for treating immune  
CC disorders, e.g., viral infections, bacterial infections, parasitic  
CC infections, neoplastic disease, autoimmunity or toxicity. The present  
CC sequence represents a spacerholder molecule peptide, which can be used in  
CC an MHC class II compound from the present invention.  
XX  
XX  
SQ Sequence 12 AA;  
Query Match 70.4%; Score 38; DB 8; Length 12;  
Best Local Similarity 66.7%; Pred. No. 8.5; Mismatches 0; Indels 0; Gaps 0;  
Matches 8; Conservative 4;  
QY 1 ASMSAASASMA 12  
|:|:|:|:|:|:  
Db 1 AAAAAAAMA 12  
RESULT 17  
ABB49912  
ID ABB49912 standard; protein; 328 AA.  
XX  
XX AC ABB49912;  
XX  
XX DT 05-FEB-2002 (first entry)  
XX  
XX DE Listeria monocytogenes protein #2616.  
XX  
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX  
XX OS Listeria monocytogenes.  
XX  
XX PN WO200177335-A2.  
XX  
XX PD 18-OCT-2001.  
XX  
XX PF 11-APR-2001; 2001WO-FR001118.  
XX  
XX PR 11-APR-2000; 2000FR-00004629.  
XX  
XX PA (INSP ) INST PASTEUR.  
XX  
XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;  
PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;  
XX  
XX WPI; 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and related  
XX polypeptides.  
XX  
PS Claim 6; SEQ ID NO 2617; 192pp; French.  
XX  
XX The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 328 AA;  
Query Match 70.4%; Score 38; DB 5; Length 328;  
Best Local Similarity 69.2%; Pred. No. 2.9e+02;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ASMSAASASMA 13  
|:|:|:|:|:|:  
Db 313 AAVSAAAAASAAA 325  
RESULT 18  
ABB71345  
ID ABB71345 standard; protein; 508 AA.  
XX  
XX AC ABB71345;  
XX  
XX DT 26-MAR-2002 (first entry)  
XX  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 40827.  
XX  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX OS Drosophila melanogaster.  
XX  
XX PN WO200171042-A2.  
XX  
XX PD 27-SEP-2001.  
XX  
XX PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX PR 23-MAR-2000; 2000US-0191637P.  
XX  
XX PR 11-JUL-2000; 2000US-00614150.  
XX  
XX PA (PEKE ) PE CORP NY.  
XX  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX N-PSDB; ABL15448.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX PS Disclosure; SEQ ID NO 40827; 21pp + Sequence Listing; English.  
XX

CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 508 AA;  
 Query Match 70.4%; Score 38; DB 4; Length 508;  
 Best Local Similarity 69.2%; Pred. No. 4.7e+02;  
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ASMSAASAAASMAA 13  
 ||:|||||:  
 DB 449 ASVAASAAAAAA 461  
 RESULT 19  
 ABR40698  
 ID ABR40698 standard; protein; 589 AA.  
 XX AC  
 XX AC ABR40698;  
 XX DT 16-MAY-2003 (first entry)  
 XX DE Glycine max oil trait related protein sequence SEQ ID NO:216.  
 XX KW Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;  
 KW receptor-like protein kinase; mitogen activated protein kinase; oil;  
 KW LIP15-like transcription factor caleosin; ATP citrate lyase; SNF1;  
 KW CKC-like transcription factor; antisense inhibition; co-suppression;  
 KW transgenic plant.  
 XX OS  
 XX OS Glycine max.  
 XX PN WO2003002751-A2.  
 XX PD 09-JAN-2003.  
 XX PF 27-JUN-2002; 2002WO-US020152.  
 XX PR 29-JUN-2001; 2001US-0301913P.  
 XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX PA (PION-) PIONEER HI-BRED INT INC.  
 XX PI Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;  
 PI Jones TJ, Kinney AJ, Klein TW, Li C, Oliveira IC, Sakai H, Shen B;  
 PI Tarczynski MC;  
 XX WPI; 2003-201509/19.  
 DR N-PSDB; ACC00733.  
 XX Novel nucleotide fragment encoding polypeptides having receptor-like  
 PT protein kinase activity, caleosin-like activity, useful for altering oil  
 PT phenotypes in plants such as sunflower, coconut, soybean, wheat and rice.  
 XX Claim 12; Page 248-250; 542pp; English.  
 XX The present invention describes an isolated nucleotide fragment (I)  
 CC comprising a nucleic acid sequence (NS) chosen from a NS encoding a  
 CC polypeptide (PP) having receptor-like protein kinase activity, mitogen  
 CC activated protein (MAP)-kinase activity, LIP15-like transcription factor  
 CC activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like  
 CC activity and CKC-like transcription factor activity. Also described: (1)  
 CC complement (II) of (I); (2) a chimeric construct (III) comprising (I) or  
 CC (II), operably linked to a regulatory sequence; (3) a plant (IV)  
 CC comprising (III) in its genome; (4) seeds (V) obtained from (IV); and (5)

CC oil obtained from (V). (I) or its part can be used in antisense  
 CC inhibition or co-suppression in a transformed plant. (III) is useful for  
 CC altering the oil phenotype in a plant such as corn, soybean, wheat, rice,  
 CC canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for  
 CC creating transgenic plants having altered lipid profiles. (I) can also be  
 CC used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to  
 CC ABR40879 represent sequences used in the exemplification of the present  
 CC invention  
 XX  
 SQ Sequence 589 AA;  
 Query Match 70.4%; Score 38; DB 6; Length 589;  
 Best Local Similarity 69.2%; Pred. No. 5.5e+02;  
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ASMSAASAAASMAA 13  
 ||:|||||:  
 DB 67 ASFRSASASSMAA 79  
 RESULT 20  
 ABR87998  
 ID ABR87998 standard; protein; 45 AA.  
 XX AC  
 XX AC ABR87998;  
 XX DT 02-JUN-2005 (first entry)  
 XX DE Rice abiotic stress responsive polypeptide SEQ ID NO:6244.  
 XX KW abiotic stress tolerance; transgenic plant; plant; cereal; agriculture.  
 XX OS  
 XX OS *Oryza sativa*.  
 XX PN WO2003008540-A2.  
 XX PD 30-JAN-2003.  
 XX PF 21-JUN-2002; 2002WO-US019668.  
 XX PR 22-JUN-2001; 2001US-0300112P.  
 PR 24-AUG-2001; 2001US-0314662P.  
 PR 26-SEP-2001; 2001US-0325277P.  
 PR 21-NOV-2001; 2001US-0332132P.  
 XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX Krepis J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;  
 PI Moughamer T, Provart N, Ricke D, Zhu T;  
 XX WPI; 2003-248011/24.  
 XX New stress-responsive nucleic acid, useful for altering the  
 PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold  
 PT stress, salt stress or osmotic stress.  
 XX Claim 1; SEQ ID NO 6244; 89pp; English.  
 PS The invention relates to novel abiotic stress responsive polynucleotides  
 CC and polypeptides. Also disclosed are vectors, expression cassettes, host  
 CC cells, and plants containing such polynucleotides. Also disclosed are  
 CC methods for using the polynucleotides and polypeptides to alter the  
 CC responsiveness of a plant to abiotic stress. The invention is useful in  
 CC agriculture. The nucleic acid is useful for determining whether a test  
 CC plant has been exposed to an abiotic stress condition. It is also useful  
 CC for selecting an agent that alters abiotic stress regulated  
 CC polynucleotide expression in a plant cell, and to identify a homolog or  
 CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid  
 CC molecule and the polypeptide encoded by it are useful in altering the  
 CC responsiveness of a plant to an abiotic stress, such as cold stress, salt  
 CC stress, osmotic stress or any of their combinations. The present sequence  
 CC is used in the exemplification of the invention  
 XX

```
SQ Sequence 45 AA;
Query Match 68.5%; Score 37; DB 7; Length 45;
Best Local Similarity 61.5%; Pred. No. 51;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ASMSAASAAASMAA 13
|::|::|::|::|
DB 18 ASUTAASASAMIA 30

RESULT 21
ABM92683
ID ABM92683 standard; protein; 170 AA.
XX AC ABM92683;
DT 02-JUN-2005 (first entry)
XX DE M. xanthus protein sequence, seq id 11882.
XX KW Transgenic plant; DNA replication; gene regulation; gene expression.
XX OS Myxococcus xanthus.
XX PN US6833447-B1.
XX PD 21-DEC-2004.
XX PF 10-JUL-2001; 2001US-00902540.
XX PR 10-JUL-2000; 2000US-0217883P.
XX PA (MONS ) MONSANTO TECHNOLOGY LLC.
XX PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX DR WPI; 2005-028716/03.
XX PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX PS Example 2; SEQ ID NO 11882; 25pp; English.
XX CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 9692-16825 represent
CC a group of 7134 Myxococcus xanthus proteins and peptides. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
XX SQ Sequence 170 AA;
Query Match 68.5%; Score 37; DB 9; Length 170;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 SMSAASAAASMAA 13
|::|::|::|::|
DB 23 SASAASAAATMVA 34

RESULT 22
ABB65204
ID ABB65204 standard; protein; 285 AA.
XX AC ABB65204;

SQ Sequence 45 AA;
Query Match 68.5%; Score 37; DB 7; Length 45;
Best Local Similarity 61.5%; Pred. No. 51;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 ASMSAASAAASMAA 13
|::|::|::|::|
DB 18 ASUTAASASAMIA 30

RESULT 21
ABM92683
ID ABM92683 standard; protein; 170 AA.
XX AC ABM92683;
DT 02-JUN-2005 (first entry)
XX DE M. xanthus protein sequence, seq id 11882.
XX KW Transgenic plant; DNA replication; gene regulation; gene expression.
XX OS Myxococcus xanthus.
XX PN US6833447-B1.
XX PD 21-DEC-2004.
XX PF 10-JUL-2001; 2001US-00902540.
XX PR 10-JUL-2000; 2000US-0217883P.
XX PA (MONS ) MONSANTO TECHNOLOGY LLC.
XX PI Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
XX DR WPI; 2005-028716/03.
XX PT New substantially purified Myxococcus xanthus nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX PS Example 2; SEQ ID NO 11882; 25pp; English.
XX CC The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 9692-16825 represent
CC a group of 7134 Myxococcus xanthus proteins and peptides. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
XX SQ Sequence 170 AA;
Query Match 68.5%; Score 37; DB 9; Length 170;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 SMSAASAAASMAA 13
|::|::|::|::|
DB 23 SASAASAAATMVA 34

RESULT 22
ABB65204
ID ABB65204 standard; protein; 285 AA.
XX AC ABB65204;

SQ Sequence 285 AA;
Query Match 68.5%; Score 37; DB 4; Length 285;
Best Local Similarity 61.5%; Pred. No. 3.6e+02;
Matches 8; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 ASMSAASAAASMAA 13
|::|::|::|::|
DB 158 AAVAASAAASAAVAA 170

RESULT 23
ABU50372
ID ABU50372 standard; protein; 375 AA.
XX AC ABU50372;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #35899.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Yersinia pestis.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
```

XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA54242.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 78296; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 375 AA;  
 Query Match 68.5%; Score 37; DB 6; Length 375;  
 Best Local Similarity 90.0%; Pred. No. 4.9e+02;  
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 SAASAAASMAA 13  
 |||||  
 DB 145 SAASVASMAA 154  
 RESULT 24  
 AAB58798  
 ID AAB58798 standard; protein; 434 AA.  
 XX  
 AC AAB58798;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 506.

XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KW neurotropic; neurprotection; antiviral; antiallergic; hepatotropic;  
 KW antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;  
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KW cardiovascular disorder; wound healing; neurological disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055173-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US005881.  
 XX  
 PR 12-MAR-1999; 99US-0124270P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-611515/58.  
 DR N-PSDB; AAF21701.  
 XX  
 PT New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention, treatment  
 PT and diagnosis of cancer, immune disorders, cardiovascular disorders and  
 PT neurological diseases.  
 XX  
 PS Claim 11; Page 940-942; 1299pp; English.  
 XX  
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive; neurotropic;  
 CC neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic;  
 CC antiinflammatory; antitumor; antiparasitic; anticonvulsant; antibacterial;  
 CC antifungal; antiparasitic and cardiant activity. The polynucleotide and  
 CC protein sequences are used in the diagnosis of cancer, particularly  
 CC breast and ovarian cancer. The nucleic acid sequences, proteins, agonists  
 CC and antagonists may also be used in the diagnosis, prevention and treatment  
 CC of immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
 CC cardiovascular disorders such as myocardial ischaemias; wound healing;  
 CC neurological diseases such as cerebral anoxia and epilepsy; and  
 CC infectious diseases  
 XX  
 SQ Sequence 434 AA;  
 Query Match 68.5%; Score 37; DB 3; Length 434;  
 Best Local Similarity 61.5%; Pred. No. 5.7e+02;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ASMSAASASMAA 13  
 |||||  
 DB 8 ASVSTAAALAA 20  
 RESULT 25  
 ADR43053  
 ID ADR43053 standard; protein; 455 AA.  
 XX  
 AC ADR43053;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX



DE IPT-like structural protein #112.  
XX  
KW abscission zone; apical meristem; Plant Growth Regulant; Gene Therapy;  
KW A6; abscission zone promoter; IPT; isopentenyl transferase.  
XX  
OS Unidentified.  
XX  
PN WO2004074442-A2.  
XX  
PD 02-SEP-2004.  
XX  
PF 13-FEB-2004; 2004WO-US0004499.  
XX  
PR 14-FEB-2003; 2003US-0447833P.  
XX  
PA (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
PI Bhat DG, Deng M, Eilers RJ, Nelson DE, Tennesen DJ;  
XX  
DR WPI; 2004-635562/61.  
XX  
XX New polynucleotides comprising a plant promoter that directs its  
PT transcription into abscission zone(s), root, pod, wall, apical meristem,  
PT or flower of plants, useful for producing a plant having increased yield  
PT or agronomic trait.  
XX  
PS Claim 21; SEQ ID NO 135; 63pp; English.  
XX  
CC The present invention relates to an isolated polynucleotide comprising a  
CC plant promoter directing transcription of an operably linked  
CC polynucleotide in at least one of the tissues selected from abscission  
CC zone, root, pod, wall, apical meristem, and flower of plants. The  
CC polynucleotides and methods are useful for producing a plant having  
CC increased yield or a trait of agronomic interest when compared to a non-  
CC transformed plant of the same genotype. The present sequence represents a  
CC IPT-like structural protein. The sequence data for this patent is not  
CC represented in the printed specification but was obtained in electronic  
CC format directly from WIPO at [www.wipo.int/pct/en/sequences/listing.htm](http://www.wipo.int/pct/en/sequences/listing.htm).  
XX  
SQ Sequence 455 AA;  
Query Match 68.5%; Score 37; DB 8; Length 455;  
Best Local Similarity 66.7%; Pred. No. 6e+02;  
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SMSAASAAASMAA 13  
Db 110 SLAAATAASIAA 121  
|:|:|:|:|:|:  
RESULT 26  
AEF15580  
ID AEF15580 standard; protein; 455 AA.  
XX  
AC AEF15580;  
XX  
DT 09-MAR-2006 (first entry)  
XX  
DE Isopentenyl transferase, SEQ ID 128.  
XX  
KW Transgenic plant; plant breeding; isopentenyl transferase; enzyme; plant.  
OS Oryza sativa.  
XX  
XX US2006010515-A1.  
PN  
PD 12-JAN-2006.  
XX  
PF 17-JUN-2005; 2005US-00156084.  
XX  
PR 23-JUN-2004; 2004US-0582357P.  
XX  
PA (HESS/) HE S S.  
XX

XX He SS;  
XX  
DR WPI; 2006-088763/09.  
XX  
XX New recombinant DNA polynucleotide comprising a cell cycle regulated  
PT promoter and encoding an isopentenyl transferase, useful for producing  
PT transgenic plants expressing cytokinin biosynthetic genes.  
XX  
XX Disclosure; SEQ ID NO 128; 21pp; English.  
XX  
XX The present invention relates to a novel recombinant DNA polynucleotide  
CC comprising a cell cycle regulated promoter, such as a cyclin promoter,  
CC that functions in plants operably linked to a DNA polynucleotide encoding  
CC an isopentenyl transferase. The present sequence is one such isopentenyl  
CC transferase. Also claimed are transgenic plants comprising the  
CC recombinant DNA polynucleotide. The recombinant polynucleotide is useful  
CC for producing transgenic plants expressing cytokinin biosynthetic genes,  
CC and also provides an improved method of producing transgenic plants with  
CC increased yield and vigor. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format directly from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
XX  
SQ Sequence 455 AA;  
Query Match 68.5%; Score 37; DB 10; Length 455;  
Best Local Similarity 66.7%; Pred. No. 6e+02;  
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
QY 2 SMSAASAAASMAA 13  
Db 110 SLAAATAASIAA 121  
|:|:|:|:|:|:  
RESULT 27  
ABO71875  
ID ABO71875 standard; protein; 493 AA.  
XX  
AC ABO71875;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polypeptide #4050.  
XX  
KW Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
DR N-PSDB; ABD05446.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 20621; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC

CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABO67826-  
 CC AB084396 represent P. aeruginosa polypeptides of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 CC  
 CC Sequence 493 AA;  
 CC  
 CC Query Match 68.5%; Score 37; DB 7; Length 493;  
 CC Best Local Similarity 76.9%; Pred. No. 6.5e+02;  
 CC Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
 |||||  
 Db 189 ASRSAASRASSAA 201

RESULT 28  
 ABU22267  
 ID ABU22267 standard; protein; 533 AA.  
 XX  
 XX  
 AC ABU22267;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #7794.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW Burkholderia mallei.  
 OS  
 OS WO200277183-A2.  
 PN  
 PN  
 PD 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 PA  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA26137.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 50191; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 CC Sequence 533 AA;

Query Match 68.5%; Score 37; DB 6; Length 533;  
 Best Local Similarity 61.5%; Pred. No. 7.1e+02;  
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ASMSAASAAASMAA 13  
 |||||  
 Db 126 AGMNAATAAAVAA 138

RESULT 29  
 ADI29007  
 ID ADI29007 standard; peptide; 13 AA.  
 XX  
 AC ADI29007;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Human MHC class II compound spaceholder molecule SEQ ID NO:3.  
 XX

XX MHC class II compound; MHC class II component; MHC class II alpha chain;  
 KW MHC class II beta chain; peptide binding groove; spaceholder molecule;  
 KW effector component; immune response; immune disorder; virucide;  
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
 KW gene therapy; viral infections; bacterial infection; parasitic infection;  
 KW neoplastic disease; autoimmunity; toxicity; human.

XX Homo sapiens.  
 OS Synthetic.  
 OS  
 XX WO2004007528-A2.  
 PN  
 XX 22-JAN-2004.  
 PD  
 XX 11-JUL-2003; 2003WO-US021767.  
 XX  
 PF 12-JUL-2002; 2002US-0395494P.  
 PR 22-JUL-2002; 2002US-0397893P.  
 PR  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 PA  
 XX Wucherpfennig KW, Seth N;  
 PI  
 XX WPI; 2004-122876/12.

XX New MHC class II compound, useful for preparing a composition for  
 PT treating immune disorders e.g. viral infections, bacterial infections,  
 PT parasitic infections, neoplastic disease, autoimmunity or toxicity.



CC Organization CD ROM supplied with the specification.  
XX  
SQ Sequence 20 AA;  
  
Query Match 66.7%; Score 36; DB 7; Length 20;  
Best Local Similarity 69.2%; Pred. No. 31;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 ASMSAASAAASMAA 13  
|||:||||:  
Db 2 ASASASSAASASA 14  
  
RESULT 32  
ADJ21502  
ID ADJ21502 standard; peptide; 20 AA.  
XX  
AC ADJ21502;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Human lung cancer-related L978P peptide fragment - SEQ ID 2064.  
XX  
KW lung tumour; cytostatic; lung cancer; human; L978P.  
XX  
OS Homo sapiens.  
XX  
PN US2003211510-A1.  
XX  
PD 13-NOV-2003.  
XX  
PF 28-OCT-2002; 2002US-00283017.  
XX  
PR 30-JUN-1999; 99US-00346492.  
PR 15-OCT-1999; 99US-00419356.  
PR 17-DEC-1999; 99US-00466867.  
PR 30-DEC-1999; 99US-00476300.  
PR 06-MAR-2000; 2000US-00519642.  
PR 22-MAR-2000; 2000US-00533077.  
PR 10-APR-2000; 2000US-00546259.  
PR 27-APR-2000; 2000US-00560406.  
PR 05-JUN-2000; 2000US-00589184.  
PR 11-JUL-2000; 2000US-00614124.  
PR 29-AUG-2000; 2000US-00651563.  
PR 08-SEP-2000; 2000US-00658824.  
PR 26-SEP-2000; 2000US-00671325.  
PR 06-OCT-2000; 2000US-00677419.  
PR 30-OCT-2000; 2000US-00702705.  
PR 13-DEC-2000; 2000US-00736457.  
PR 03-MAY-2001; 2001US-00849626.  
PR 10-JUL-2001; 2001US-00902941.  
PR 29-OCT-2001; 2001US-00017754.  
PR 28-MAR-2002; 2002US-00113872.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Henderson RA, Wang T, Watanabe Y, Kalos MD, Sleath PR;  
PI Johnson JC, Retter MW, Durham M, Carter D, Fanger GR, Vedvick TS;  
PI Bangur CS, McNabb A;  
XX  
DR WPI; 2004-167010/16.  
XX  
XX Novel polynucleotide encoding lung tumor polypeptides, useful for  
PT diagnosing, preventing and treating cancer e.g. lung cancer.  
PS  
PS Example 37; SEQ ID NO 2064; 99pp; English.  
XX  
CC The invention relates to a novel isolated polynucleotide comprising a  
CC sequence chosen from any one of 40 lung tumour polypeptides or its  
CC complements, fragments or degenerate variants. The method of the  
CC invention has cytostatic applications and may be useful for detecting and  
CC treating lung cancer in a patient, as well as for inhibiting the  
CC development of lung cancer in a patient via incubating CD4+ and/or CD8+ T

CC cells isolated from a patient with at least one component chosen from a  
CC polypeptide, polynucleotide or antigen presenting cell (APC) of the  
CC invention and administering an effective amount of the proliferated T  
CC cells to the patient. The current sequence is that of the human lung  
CC cancer-related peptide of the invention. The current sequence is not  
CC shown in the specification per se but is available on the USPTO web-site  
CC <http://seqdata.uspto.gov/sequence.html?DocID=20030211510>.  
XX  
SQ Sequence 20 AA;  
  
Query Match 66.7%; Score 36; DB 8; Length 20;  
Best Local Similarity 69.2%; Pred. No. 31;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 ASMSAASAAASMAA 13  
|||:||||:  
Db 7 ASASASSAASASA 19  
  
RESULT 33  
ADJ21503  
ID ADJ21503 standard; peptide; 20 AA.  
XX  
AC ADJ21503;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Human lung cancer-related L978P peptide fragment - SEQ ID 2065.  
XX  
KW lung tumour; cytostatic; lung cancer; human; L978P.  
XX  
OS Homo sapiens.  
XX  
PN US2003211510-A1.  
XX  
PD 13-NOV-2003.  
XX  
PF 28-OCT-2002; 2002US-00283017.  
XX  
PR 30-JUN-1999; 99US-00346492.  
PR 15-OCT-1999; 99US-00419356.  
PR 17-DEC-1999; 99US-00466867.  
PR 30-DEC-1999; 99US-00476300.  
PR 06-MAR-2000; 2000US-00519642.  
PR 22-MAR-2000; 2000US-00533077.  
PR 10-APR-2000; 2000US-00546259.  
PR 27-APR-2000; 2000US-00560406.  
PR 05-JUN-2000; 2000US-00589184.  
PR 11-JUL-2000; 2000US-00614124.  
PR 29-AUG-2000; 2000US-00651563.  
PR 08-SEP-2000; 2000US-00658824.  
PR 26-SEP-2000; 2000US-00671325.  
PR 06-OCT-2000; 2000US-00677419.  
PR 30-OCT-2000; 2000US-00702705.  
PR 13-DEC-2000; 2000US-00736457.  
PR 03-MAY-2001; 2001US-00849626.  
PR 10-JUL-2001; 2001US-00902941.  
PR 29-OCT-2001; 2001US-00017754.  
PR 28-MAR-2002; 2002US-00113872.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Henderson RA, Wang T, Watanabe Y, Kalos MD, Sleath PR;  
PI Johnson JC, Retter MW, Durham M, Carter D, Fanger GR, Vedvick TS;  
PI Bangur CS, McNabb A;  
XX  
DR WPI; 2004-167010/16.  
XX  
XX Novel polynucleotide encoding lung tumor polypeptides, useful for  
PT diagnosing, preventing and treating cancer e.g. lung cancer.  
PS  
PS Example 37; SEQ ID NO 2065; 99pp; English.  
XX  
XX

CC The invention relates to a novel isolated polynucleotide comprising a  
CC sequence chosen from any one of 40 lung tumour polypeptides or its  
CC complements, fragments or degenerate variants. The method of the  
CC invention has cytostatic applications and may be useful for detecting and  
CC treating lung cancer in a patient, as well as for inhibiting the  
CC development of lung cancer in a patient via incubating CD4+ and/or CD8+ T  
CC cells isolated from a patient with at least one component chosen from a  
CC polypeptide, polynucleotide or antigen presenting cell (APC) of the  
CC invention and administering an effective amount of the proliferated T  
CC cells to the patient. The current sequence is that of the human lung  
CC cancer-related peptide of the invention. The current sequence is not  
CC shown in the specification per se but is available on the USPTO web-site  
CC <http://seqdata.uspto.gov/sequence.html?DocID=20030211510>.  
XX  
SQ Sequence 20 AA;  
Query Match 66.7%; Score 36; DB 8; Length 20;  
Best Local Similarity 69.2%; Pred. No. 31;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 ASMSAASAAASMAA 13  
|||:|||||:  
Db 2 ASASASSAASASA 14  
RESULT 34  
ABB70649  
ID ABB70649 standard; protein; 148 AA.  
XX  
AC ABB70649;  
XX  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 38739.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
XX  
XX 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX  
XX N-PSDB; ABL14752.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX  
XX Disclosure; SEQ ID NO 38739; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](ftp://wipo.int/pub/published_pct_sequences)

SQ Sequence 148 AA;  
Query Match 66.7%; Score 36; DB 4; Length 148;  
Best Local Similarity 76.9%; Pred. No. 2.6e+02;  
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 ASMSAASAAASMAA 13  
|||:|||||:  
Db 50 ASASAAVAASAAA 62  
RESULT 35  
ADT59595  
ID ADT59595 standard; protein; 250 AA.  
XX  
XX AC ADT59595;  
XX  
XX 13-JAN-2005 (first entry)  
XX  
XX Plant polypeptide, SEQ ID 9672.  
XX  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
XX disease resistance; galactomannan production; plant growth regulator;  
XX heat tolerance; herbicide tolerance; lignin production;  
XX extreme osmotic condition tolerance; pathogens resistance;  
XX pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
XX Viridiplantae.  
XX  
XX US2004216190-A1.  
XX  
XX 28-OCT-2004.  
XX  
XX 18-DEC-2003; 2003US-00739930.  
XX  
XX 28-APR-2003; 2003US-00424599.  
XX  
XX 28-APR-2003; 2003US-00425115.  
XX  
XX (KOVA/) KOVALIC D K.  
XX  
XX Kovalic DK;  
XX  
XX WPI; 2004-757369/74.  
XX  
XX New recombinant DNA constructs useful in the field of biochemistry and  
XX genetics, and in particular for producing transgenic plants with improved  
XX biological characteristics.  
XX  
XX Claim 2; SEQ ID NO 9672; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
XX polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
XX 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
XX (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
XX Arabidopsis, wheat and rape but the specification does not indicate which  
XX sequences is derived from which organism. Also included is a method of  
XX producing a plant having an improved property, comprising transforming a  
XX plant with a recombinant DNA construct comprising a promoter region  
XX functional in a plant cell operably joined to a polynucleotide encoding a  
XX polypeptide associated with the property, and growing the transformed  
XX plant. The property is selected from improving plant cold tolerance, for  
XX manipulating growth rate in plant cells by modification of the cell cycle  
XX pathway, for improving plant drought tolerance, for providing increased  
XX resistance to plant disease, for galactomannan production, for production  
XX of plant growth regulators, for improving plant heat tolerance, for  
XX improving plant tolerance to herbicides, for increasing the rate of  
XX homologous recombination in plants, for lignin production, for improving  
XX plant tolerance to extreme osmotic conditions, for improving plant  
XX tolerance to pathogens or pests, for yield improvement by modification of  
XX photosynthesis, for modifying seed oil yield and/or content, for  
XX photosynthesis, for modifying seed oil yield and/or content, for  
XX modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
XX and for yield improvement by providing improved plant growth and

CC development under at least one stress condition. The polynucleotide may  
 CC also encode a plant transcription factor. The methods and compositions of  
 CC the present invention are useful in the field of biochemistry and  
 CC genetics, in particular for producing transgenic plants with improved  
 CC biological characteristics such as increased yield, improved nitrogen  
 CC flow, increasing plant tolerance to cold or heat, improving plant  
 CC tolerance to extreme osmotic and drought conditions, and improving plant  
 CC tolerance to plant pests or pathogens. They can also be used in physical  
 CC arrays of molecules, plant breeding markers, computer-based storage and  
 CC analysis systems. The present sequence is one of the 5544 plant protein  
 CC sequences of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
 CC  
 XX  
 SQ Sequence 250 AA;

Query Match 66.7%; Score 36; DB 8; Length 250;  
 Best Local Similarity 80.0%; Pred. No. 4.6e+02;  
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 MSAASAASMA 12  
 |:::|||||  
 Db 1 MAAAAASMA 10

RESULT 36  
 AB083252  
 ID AB083252 standard; protein; 268 AA.  
 XX  
 AC AB083252;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Pseudomonas aeruginosa polypeptide #15427.  
 XX  
 KW Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX Pseudomonas aeruginosa.  
 OS  
 XX  
 PN US6551795-B1.  
 XX  
 PD 22-APR-2003.  
 XX  
 PF 18-FEB-1999; 99US-00252991.  
 XX  
 PR 18-FEB-1999; 98US-0074788P.  
 PR 27-JUL-1998; 98US-0094190P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
 XX  
 DR WPI; 2003-615309/58.  
 DR N-PSDB; ABD16823.  
 XX  
 PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 XX  
 PS Disclosure; SEQ ID NO 31998; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences

CC of Pseudomonas species using biochip technology. Sequences AB067826-  
 CC AB04396 represent P. aeruginosa polypeptides of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 XX  
 SQ Sequence 268 AA;

Query Match 66.7%; Score 36; DB 7; Length 268;  
 Best Local Similarity 81.8%; Pred. No. 4.9e+02;  
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 MSAASAASMA 13  
 ||| |||||  
 Db 152 MSPARAASMA 162

RESULT 37  
 ADR98917  
 ID ADR98917 standard; protein; 278 AA.  
 XX  
 AC ADR98917;  
 XX  
 DT 18-NOV-2004 (first entry)

XX  
 DE Lung specific gene splice variant encoded protein #89.  
 XX  
 KW cytostatic; gene therapy; vaccine; lung; diagnosis; cancer;  
 KW non-cancerous lung disease; lung tissue; antagonist; gene therapy;  
 KW transgenic animal; splice variant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004074430-A2.  
 XX  
 PD 02-SEP-2004.

XX  
 PF 08-DEC-2003; 2003WO-US038896.  
 XX  
 PR 06-DEC-2002; 2002US-0431307P.  
 PR 06-DEC-2002; 2002US-0431510P.  
 PR 06-DEC-2002; 2002US-0431516P.  
 XX  
 PA (DIAD-) DIADEXUS INC.

XX  
 PI Macina RA, Turner LR, Sun Y, Liu S;  
 XX  
 DR WPI; 2004-635553/61.  
 DR N-PSDB; ADR98780.

XX  
 PT New isolated human lung specific nucleic acid molecule, useful for  
 PT identifying, diagnosing, monitoring, staging, imaging and treating lung  
 PT cancer and non-cancerous diseases of the lung.  
 XX  
 PS Claim 1; SEQ ID NO 202; 542pp; English.

XX The invention relates to a new isolated lung specific nucleic acid  
 CC molecule (I) comprising any of 113 fully defined nucleotide sequences  
 CC given in the specification, their encoded protein sequences, sequences  
 CC selectively hybridizing to the nucleotide sequences or a sequence having  
 CC at least 60% identity to the nucleotide sequences. The methods and  
 CC compositions of the present invention are useful for identifying,  
 CC diagnosing, monitoring, staging, imaging and treating lung cancer and non  
 CC -cancerous diseases of the lung. They are also used for identifying lung  
 CC tissue, monitoring and identifying and/or designing antagonists of the  
 CC polypeptide of the invention, gene therapy, production of transgenic  
 CC animals and production of engineered lung tissue for treatment and  
 CC research. Lung specific genes (LSGs) were identified by a systematic  
 CC analysis of gene expression data in the LIFESEQ Gold database using the  
 CC data mining software package candidate lead automatic search program  
 CC (CLASP). Genes were grouped into gene bins where each bin is a cluster of  
 CC sequences grouped together where they share a common contig.  
 CC Differentially expressed tissue-specific genes were selected based on the

CC percentage level in the targeted tissue versus all the other tissues. The  
CC expression levels for each gene in libraries of normal tissues or non-  
CC tumour tissues from cancer patients were compared with the expression  
CC levels in tissue libraries associated with tumour or disease. This  
CC sequence represents a protein of the invention.

XX Sequence 278 AA;

Query Match 66.7%; Score 36; DB 8; Length 278;  
Best Local Similarity 61.5%; Pred. No. 5.1e+02;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
DB 1 AAAAAAASMAA 13

RESULT 38  
ABB66932  
ID ABB66932 standard; protein; 327 AA.  
XX AC ABB66932;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 27588.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL11035.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.

XX Disclosure; SEQ ID NO 27588; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
XX ABBS72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 327 AA;

Query Match 66.7%; Score 36; DB 4; Length 327;  
Best Local Similarity 69.2%; Pred. No. 6.1e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13  
||| ||| ||| |||

Db 214 ASTSAAAAA 226

RESULT 39

AAB94055

ID AAB94055 standard; protein; 327 AA.

XX AC AAB94055;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:14226.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
XX length cDNAs defined in the specification, and for the detection and/or  
XX diagnosis of the abnormality of the proteins encoded by the full-length  
XX cDNAs.

XX Claim 8; SEQ ID NO 14226; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesising 5602 full-  
XX length cDNAs defined in the specification. Where a primer set comprises:  
XX (a) an oligo-dT primer and an oligonucleotide complementary to the  
XX complementary strand of a polynucleotide which comprises one of the 5602  
XX nucleotide sequences defined in the specification, where the  
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
XX of an oligonucleotide comprising a sequence complementary to the  
XX complementary strand of a polynucleotide which comprises a 5'-end  
XX sequence and an oligonucleotide comprising a sequence complementary to a  
XX polynucleotide which comprises a 3'-end sequence, where the  
XX oligonucleotide comprises at least 15 nucleotides and the combination of  
XX the 5'-end sequence/3'-end sequence is selected from those defined in the  
XX specification. The primer sets can be used in antisense therapy and in  
XX gene therapy. The primers are useful for synthesising polynucleotides,  
XX particularly full-length cDNAs. The primers are also useful for the  
XX detection and/or diagnosis of the abnormality of the proteins encoded by  
XX the full-length cDNAs. The primers allow obtaining of the full-length  
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893  
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent  
XX oligonucleotides, all of which are used in the exemplification of the  
XX present invention

XX Sequence 327 AA;

Query Match 66.7%; Score 36; DB 4; Length 327;  
Best Local Similarity 69.2%; Pred. No. 6.1e+02;  
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 ASMSAASASMAA 13

Db ||||:|||| ||||:|  
2 ASMAAIAASRSA 14

RESULT 40

AAM93990  
ID AAM93990 standard; protein; 327 AA.

XX AC AAM93990;

XX DT 13-NOV-2001 (first entry)

XX DE Human stomach cancer expressed polypeptide SEQ ID NO 49.

XX KW Human; stomach cancer; marker; screening; micro-metastasis;  
peritoneal dissemination.

XX OS Homo sapiens.

PN WO200109317-A1.

XX PD 08-FEB-2001.

XX PF 28-JUL-2000; 2000WO-JP005063.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 18-OCT-1999; 99US-0159590P.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 17-FEB-2000; 2000US-0183322P.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;  
PI Kodama T, Midorikawa Y;

DR WPI; 2001-570287/64.

DR N-PSDB; AAI93866.

XX PT New Stomach cancer-associated genes, useful as markers in blood tests for  
screening for the early stages of the disease.

XX PS Claim 1; Page 119-120; 242pp; Japanese.

XX CC The invention relates to stomach cancer-expressed genes (AAI93842-  
AAI93917) and the encoded proteins (AAM93967-AAM94039). The genes can be  
used as markers in blood tests for screening for the early stages of the  
disease. The proteins and peptides can be used as targets for screening  
for compounds to treat the disease. They can also be used for predicting  
micro-metastases. The gene can predict peritoneal dissemination

SQ Sequence 327 AA;

Query Match 66.7%; Score 36; DB 4; Length 327;

Best Local Similarity 69.2%; Pred. No. 6.1e+02;

Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 1 ASMSAASASMAA 13  
|||:|||||:|

Db 2 ASMAAIAASRSA 14  
||||:|||||:|

Search completed: September 9, 2006, 22:48:03

Job time : 103.418 secs



GenCore version 5.1.9

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:48:41 ; Search time 14.8101 Seconds  
(without alignments)  
84.457 Million cell updates/sec

Title: US-10-617-568-3

Perfect score: 53

Sequence: 1 AAAAAAAAAAAAA 13

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR\_80.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	92.5	377	2 T04213	heat shock transcr
2	49	92.5	530	2 I38558	Mi-2 autoantigen 2
3	49	92.5	577	2 S72227	finger protein sob
4	49	92.5	892	2 T09071	SH3 domains-contai
5	49	92.5	1533	2 A46221	abdominal segment
6	49	92.5	2038	2 A43742	female sterile hom
7	48	90.6	451	2 D88395	protein P53A3.6 li
8	47	88.7	150	2 T12547	hypothetical prote
9	47	88.7	220	2 JCS594	ribosomal protein
10	47	88.7	233	2 S11563	probable MASH-2 pr
11	47	88.7	238	2 A48279	achaeete scute prot
12	47	88.7	289	2 A43562	homeotic protein H
13	47	88.7	305	2 I57039	genomic screen hom
14	47	88.7	314	2 JCS5273	paired type homeob
15	47	88.7	323	2 S16318	homeotic protein H
16	47	88.7	328	2 A81781	hypothetical prote
17	47	88.7	331	2 B47236	zinc-finger protei
18	47	88.7	333	2 A39085	homeotic protein E
19	47	88.7	334	2 G02409	protein kinase C-b
20	47	88.7	364	2 T048188	gene NKx6.1 protei
21	47	88.7	374	2 T03875	probable homeobox
22	47	88.7	375	2 T03874	probable homeobox
23	47	88.7	378	2 A44443	basic helix-loop-h
24	47	88.7	401	2 A48423	engrailed homeodom
25	47	88.7	403	2 A53662	homeotic protein H
26	47	88.7	420	2 T14911	bZIP DNA-binding p
27	47	88.7	443	1 I38239	transcription fact
28	47	88.7	475	2 A43915	homeotic protein e
29	47	88.7	477	2 A47236	zinc-finger protei

30	47	88.7	482	2 JC7583	basic helix-loop-h
31	47	88.7	490	2 T09084	phosphatidylinosit
32	47	88.7	494	2 A42170	zinc finger protei
33	47	88.7	495	1 S31223	transcription fact
34	47	88.7	497	2 JC5076	myc-associated zin
35	47	88.7	513	2 A48233	polyomavirus enhan
36	47	88.7	537	2 A55929	zinc finger protei
37	47	88.7	552	1 WJFFEN	homeotic protein e
38	47	88.7	627	2 T02610	probable YME1 ATP-
39	47	88.7	644	2 S39356	transcription fact
40	47	88.7	779	1 S40382	box A-binding fact
41	47	88.7	796	2 JC7555	Cl4orf4 protein -
42	47	88.7	805	2 T49385	hypothetical prote
43	47	88.7	806	2 T13690	hypothetical prote
44	47	88.7	828	2 C88402	protein H05C05.1 [
45	47	88.7	829	2 A34692	ecdysone-induced p
46	47	88.7	873	2 B53225	ecdysone-induced p
47	47	88.7	883	2 S04722	puff 74E protein -
48	47	88.7	1028	2 A56038	DNA-binding protei
49	47	88.7	1065	2 T13230	dachshund isoform
50	47	88.7	1072	2 T13232	dachshund protein
51	47	88.7	1074	2 T13229	dachshund protein
52	47	88.7	1081	2 T13231	dachshund protein
53	47	88.7	1180	2 S69205	stripe a/b protein
54	47	88.7	1205	2 A55015	bumetanide-sensiti
55	47	88.7	1212	2 A57187	bumetanide-sensiti
56	47	88.7	1213	2 S16356	ovo protein - frui
57	47	88.7	1355	2 S40022	spalt protein - fr
58	47	88.7	1506	2 T51900	related to PAN2 pr
59	47	88.7	1668	2 T13748	sex comb protein -
60	47	88.7	1891	2 T13594	hypothetical prote
61	47	88.7	2639	2 T31328	fibroin - Chinese
62	46	86.8	470	2 S33639	finger protein esc
63	46	86.8	2175	1 S03170	homeotic protein c
64	45	84.9	306	1 S59863	polyA binding prot
65	45	84.9	337	2 S06956	segmentation prote
66	45	84.9	392	2 B48423	homeotic protein e
67	45	84.9	748	2 T49633	glucan 1,4-alpha-g
68	45	84.9	1077	2 A44067	serine-rich protei
69	45	84.9	1741	2 T13610	parallel sister ch
70	44	83.0	219	1 TE8EG6	tegument protein -
71	44	83.0	219	1 TE8EKA	DNA-binding protei
72	44	83.0	448	2 S17370	homeotic protein B
73	44	83.0	604	2 A39369	Om(1D) protein - f
74	44	83.0	606	2 S13367	homeotic protein B
75	44	83.0	640	2 A41726	homeotic protein B
76	44	83.0	642	2 S27806	castor protein - f
77	44	83.0	799	2 JH0797	neurogenic protein
78	44	83.0	832	2 A31246	neurogenic repetit
79	44	83.0	833	2 S19087	gene Delta protein
80	44	83.0	880	2 S00670	hypothetical prote
81	44	83.0	1166	2 H86341	neurogenic locus m
82	44	83.0	1596	2 A33106	gene mastermind pr
83	44	83.0	1655	2 T13998	notch protein - fr
84	44	83.0	2703	1 A24420	eyelid - fruit fly
85	44	83.0	2715	2 T31049	hypothetical prote
86	44	83.0	2957	2 T33152	BiR repeat contain
87	44	83.0	4845	2 T31067	microtubule-associ
88	44	83.0	5327	2 T13564	homeotic protein u
89	43	81.1	40	2 S58853	antifreeze protein
90	43	81.1	85	1 FDFL4W	antifreeze protein
91	43	81.1	91	2 A22592	antifreeze protein
92	43	81.1	97	2 S02376	sex-determining pr
93	43	81.1	172	2 S35568	cgcr-1 protein - C
94	43	81.1	205	2 S19114	achaeete-scute locu
95	43	81.1	231	2 S28186	hypothetical prote
96	43	81.1	308	2 A44495	transcription repr
97	43	81.1	328	2 A21405	segmentation prote
98	43	81.1	376	2 A26066	transcription fact
99	43	81.1	425	1 JH0710	paired box protein
100	43	81.1	425	2 A38153	

ALIGNMENTS

RESULT 1  
T04213  
heat shock transcription factor homolog T5C23.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T04213  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z15261  
A:Accession: T04213  
A:Molecule type: DNA  
A:Residues: 1-377 <BEV>  
A:Cross-references: UNIPROT:Q9T0D3; UNIPARC:UPI000000117B; EMBL:AL049500  
A:Experimental source: cultivar Columbia; BAC clone T5C23  
C:Genetics:  
A:Map position: 4  
A:Introns: 118/3  
A:Note: T5C23.90  
F:60-155/Domain: HSF DNA-binding domain homology <HSF>  
  
Query Match 92.5%; Score 49; DB 2; Length 377;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 2 AAAAAAAAAAAAAA 13  
|||:|||||  
Db 155 AAAAAAAAAAAAAA 166  
  
RESULT 2  
I38558  
Mi-2 autoantigen 240 kDa protein - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 05-Nov-1999  
C:Accession: I38558  
R:Ge, Q.; Nilasena, D.S.; O'Brien, C.A.; Frank, M.B.; Targoff, I.N.  
J. Clin. Invest. 96, 1730-1737, 1995  
A:Title: Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen  
A:Reference number: I38558; MUID:96013633; PMID:7560064  
A:Accession: I38558  
A:Molecule type: mRNA  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Residues: 1-530 <RES>  
A:Cross-references: UNIPARC:UPI000016A09A; EMBL:U08379; NID:g761717; PIDN:AAC50228.1; PI  
  
Query Match 92.5%; Score 49; DB 2; Length 530;  
Best Local Similarity 92.3%; Pred. No. 15;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
|||:|||||  
Db 87 AAAAAAAAAAAAAA 99  
  
RESULT 3  
S72227  
finger protein sob - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 04-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
C:Accession: S72227  
R:Hart, M.C.; Wang, L.; Coulter, D.E.  
Genetics 144, 171-182, 1996  
A:Title: Comparison of the structure and expression of odd-skipped and two related genes  
A:Reference number: S72227; MUID:97032935; PMID:8878683  
A:Accession: S72227  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-577 <HAR>  
A:Cross-references: UNIPROT:Q24571; UNIPARC:UPI000007A28D; EMBL:U62004; NID:g1480193; PI  
C:Genetics:

A:Gene: sob  
A:Cross-references: FlyBase:FBgn0004892  
C:Keywords: zinc finger  
  
Query Match 92.5%; Score 49; DB 2; Length 577;  
Best Local Similarity 92.3%; Pred. No. 16;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
|||:|||||  
Db 273 AAAAAAAAAAAAAA 285  
  
RESULT 4  
T09071  
SH3 domains-containing protein POSH - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: T09071  
R:Tapon, N.; Nagata, K.; Lamarche, N.; Hall, A.  
EMBO J. 17, 1395-1404, 1998  
A:Reference number: Z16552; MUID:98151363; PMID:9482736  
A:Accession: T09071  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-892 <TAP>  
A:Cross-references: UNIPROT:O70254; UNIPARC:UPI0000027BF6; EMBL:AF030131; NID:g3002587;  
A:Experimental source: cell line Ras-transformed NIH 3T3 cells  
A:Note: activates JNK/SAPK cascade; Rac-binding protein  
C:Genetics:  
A:Gene: POSH  
C:Keywords: signal transduction  
F:8-58/Domain: RING finger homology <RRN>  
  
Query Match 92.5%; Score 49; DB 2; Length 892;  
Best Local Similarity 92.3%; Pred. No. 22;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
|||:|||||  
Db 419 AAAAAAAAAAAAAA 431  
  
RESULT 5  
A46221  
abdominal segment formation protein pumilio - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C>Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: A46221; S22026  
R:Barker, D.D.; Wang, C.; Moore, J.; Dickinson, L.K.; Lehmann, R.  
Genes Dev. 6, 2312-2326, 1992  
A:Title: Pumilio is essential for function but not for distribution of the Drosophila ab  
A:Reference number: A46221; MUID:93093466; PMID:1459455  
A:Contents: embryo  
A:Accession: A46221  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1533 <BAR>  
A:Cross-references: UNIPROT:P25822; UNIPARC:UPI000016BD14; GB:L07943; NID:g158190; PID:9  
R:Macdonald, P.M.  
submitted to the EMBL Data Library, October 1991  
A:Reference number: S22026  
A:Accession: S22026  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-361, 'A', '363-1102, 'R', '1104-1405, 'KN', '1408-1495, 'V', '1497-1518, 'S', '1520-1533  
A:Cross-references: UNIPARC:UPI00001329FD; EMBL:X62589; NID:g8393; PID:g8394  
C:Genetics:  
A:Gene: FlyBase:pum  
A:Cross-references: FlyBase:FBgn0003165  
  
Query Match 92.5%; Score 49; DB 2; Length 1533;

Best Local Similarity 92.3%; Pred. No. 33;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
||:|||||  
Db 1050 AAAAAAAAAAAAAA 1062

RESULT 6  
A43742  
female sterile homeotic protein, 205K - fruit fly (*Drosophila melanogaster*)  
N;Alternate names: membrane protein fsh, 205K  
N;Contains: female sterile homeotic protein, 110K  
C;Species: *Drosophila melanogaster*  
C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 31-Dec-2004  
C;Accession: A43742; B43742  
R;Haynes, S.R.; Mozer, B.A.; Bhatia-Dey, N.; Dawid, I.B.  
Dev. Biol. 134, 246-257, 1989  
A;Title: The *Drosophila* fsh locus, a maternal effect homeotic gene, encodes apparent mem  
A;Reference number: A43742; MUID:89276730; PMID:2567251  
A;Accession: A43742  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-2038 <HAY>  
A;Cross-references: UNIPROT:P13709; UNIPARC:UPI000012AC6C; EMBL:M23221; NID:g157452; PID  
A;Accession: B43742  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1106 <HA2>  
A;Cross-references: UNIPARC:UPI000017A10F; EMBL:M23222  
C;Genetics:  
A;Gene: fsh  
A;Cross-references: FlyBase:FBgn0004656  
C;Keywords: alternative splicing; transmembrane protein  
F;1-2038/Product: female sterile homeotic protein, 205K #status predicted <NA2>  
F;1-1106/Product: female sterile homeotic protein, 110K #status predicted <NA1>  
F;59-116/Domain: bromodomain homology <BRO1>  
F;503-560/Domain: bromodomain homology <BRO2>

Query Match 92.5%; Score 49; DB 2; Length 2038;  
Best Local Similarity 92.3%; Pred. No. 41;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
||:|||||  
Db 330 AAAAAAAAAAAAAA 342

RESULT 7  
D88395  
protein F53A3.6 [imported] - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C;Accession: D88395  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biolog  
A;Reference number: A75000; MUID:99069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A;Accession: D88395  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-451 <STO>  
A;Cross-references: UNIPROT:O17222; UNIPARC:UPI000017A5BC; GB:chr\_III; PIDN:AAB70991.1;  
C;Genetics:  
A;Gene: F53A3.6  
A;Map position: 3

Query Match 90.6%; Score 48; DB 2; Length 451;  
Best Local Similarity 92.3%; Pred. No. 18;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
||:|||||  
Db 303 AAAAAAAAAAAAAA 315

RESULT 8  
T12547  
hypothetical protein DKFZp586E1621.1 - human (fragment)  
C;Species: *Homo sapiens* (man)  
C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C;Accession: T12547  
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A;Reference number: Z17528  
A;Accession: T12547  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-150 <OTT>  
A;Cross-references: UNIPROT:Q9Y4M1; UNIPARC:UPI0000071DDD; EMBL:AL080235  
A;Experimental source: adult uterus; clone DKFZp586E1621  
C;Genetics:  
A;Note: DKFZp586E1621.1

Query Match 88.7%; Score 47; DB 2; Length 150;  
Best Local Similarity 92.3%; Pred. No. 10;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
||:|||||  
Db 129 AAAAAAAAAAAAAA 141

RESULT 9  
JCS954  
ribosomal protein L14 - human  
C;Species: *Homo sapiens* (man)  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 02-Sep-2000  
C;Accession: JCS954  
R;Tanaka, M.; Tanaka, T.; Harata, M.; Suzuki, T.; Mitsuui, Y.  
Biochem. Biophys. Res. Commun. 243, 531-537, 1998  
A;Title: Triplet repeat-containing ribosomal protein L14 gene in immortalized human end  
A;Reference number: JCS954; MUID:98153799; PMID:9480843  
A;Accession: JCS954  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-220 <TAN>  
A;Cross-references: UNIPARC:UPI000004E622; DDBJ:D87735; NID:g1620021; PIDN:BAAL3443.1;  
C;Superfamily: rat ribosomal protein L14

Query Match 88.7%; Score 47; DB 2; Length 220;  
Best Local Similarity 92.3%; Pred. No. 14;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
||:|||||  
Db 150 AAAAAAAAAAAAAA 162

RESULT 10  
S11563  
probable MASH-2 protein - rat  
C;Species: *Rattus norvegicus* (Norway rat)  
C;Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C;Accession: S11563  
R;Johnson, J.E.; Birren, S.J.; Anderson, D.J.  
Nature 346, 858-861, 1990  
A;Title: Two rat homologues of *Drosophila* achaete-scute specifically expressed in neuro  
A;Reference number: S11562; MUID:90363294; PMID:2392153  
A;Accession: S11563  
A;Status: preliminary; nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-233 <JOH>  
A;Cross-references: UNIPROT:P19359; UNIPARC:UPI0000047567; EMBL:X53725; NID:g56630; PID

C;Keywords: DNA binding

Query Match 88.7%; Score 47; DB 2; Length 233;  
Best Local Similarity 92.3%; Pred. No. 14;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 32 AAAAAAAAAAAAAA 44

#### RESULT 11

A48279  
achaete scute protein - human  
C;Species: Homo sapiens (man)  
C;Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 09-Jul-2004  
C;Accession: A48279  
R;Ball, D.W.; Azzoli, C.G.; Baylin, S.B.; Chi, D.; Dou, S.; Donis-Keller, H.; Kumaraswami, P.; Natl. Acad. Sci. U.S.A. 90, 5648-5652, 1993  
A;Title: Identification of a human achaete-scute homolog highly expressed in neuroendocrine cells  
A;Reference number: A48279; MUID:93296195; PMID:8390674  
A;Accession: A48279  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-238 <RES>  
A;Cross-references: UNIPROT:P50553; UNIPARC:UPI0000047566; GB:L08424; NID:G306459; PID:G306459

Query Match 88.7%; Score 47; DB 2; Length 238;  
Best Local Similarity 92.3%; Pred. No. 15;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 35 AAAAAAAAAAAAAA 47

#### RESULT 12

A43562  
homeotic protein Hox D8 - mouse  
N;Alternate names: homeotic protein Hox 4.3  
C;Species: Mus musculus (house mouse)  
C;Date: 11-Dec-1992 #sequence\_revision 11-Dec-1992 #text\_change 09-Jul-2004  
C;Accession: A43562  
R;Izpisua-Belmonte, J.C.; Dolle, P.; Renucci, A.; Zappavigna, V.; Falkenstein, H.; Duboulet, J.L.; Development 110, 733-745, 1990  
A;Title: Primary structure and embryonic expression pattern of the mouse Hox-4.3 homeobox gene  
A;Reference number: A43562; MUID:91209232; PMID:1982431  
A;Accession: A43562  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-289 <IZP>  
A;Cross-references: UNIPROT:P23463; UNIPARC:UPI0000177DA6  
C;Superfamily: homeotic protein Hox A7; homeobox homology  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;196-252/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 289;  
Best Local Similarity 92.3%; Pred. No. 17;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 15 AAAAAAAAAAAAAA 27

#### RESULT 13

I57039  
genomic screen homeobox protein 2 - mouse  
C;Species: Mus sp. (mouse)  
C;Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 05-Oct-2004  
C;Accession: I57039  
R;Hsieh-Li, H.M.; Witte, D.P.; Szucsik, J.C.; Weinstein, M.; Li, H.; Potter, S.S. Mech. Dev. 50, 177-186, 1995

A;Title: Gsh-2, a murine homeobox gene expressed in the developing brain.  
A;Reference number: I57039; MUID:95344993; PMID:7619729

A;Accession: I57039  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: mRNA  
A;Residues: 1-305 <RES>  
A;Cross-references: UNIPARC:UPI0000024B69; GB:S79041; NID:G1042008; PIDN:AAB34947.1; PIDN:AAB34947.1

C;Genetics:  
A;Gene: Gsh-2

C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;204-260/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 305;  
Best Local Similarity 92.3%; Pred. No. 18;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 147 AAAAAAAAAAAAAA 159

#### RESULT 14

JC5273  
paired type homeobox protein, NBP - human  
C;Species: Homo sapiens (man)  
C;Date: 16-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 05-Oct-2004  
C;Accession: JC5273  
R;Yokoyama, M.; Nishi, Y.; Yoshii, J.; Okubo, K.; Matsubara, K. DNA Res. 3, 311-320, 1996  
A;Title: Identification and cloning of neuroblastoma-specific and nerve tissue-specific homeobox genes  
A;Reference number: JC5272; MUID:97191543; PMID:9039501  
A;Contents: neuroblastoma cell

A;Accession: JC5273  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA

A;Residues: 1-314 <YOK>  
A;Cross-references: UNIPROT:Q99453; UNIPARC:UPI0000131D16; DBJ:D82344; NID:G1841337; PIDN:G1841337

C;Comment: This protein is a transcriptional repressor involved in regulating gene expression  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;99-155/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 314;  
Best Local Similarity 92.3%; Pred. No. 18;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 241 AAAAAAAAAAAAAA 253

#### RESULT 15

S16318  
homeotic protein Hox 4.6 - mouse  
N;Alternate names: homeotic protein Hoxd-11  
C;Species: Mus musculus (house mouse)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Oct-2004  
C;Accession: S16318; S40403; S57443  
R;Izpisua-Belmonte, J.C.; Falkenstein, H.; Dolle, P.; Renucci, A.; Duboule, D. EMBO J. 10, 2279-2289, 1991  
A;Title: Murine genes related to the Drosophila AbdB homeotic gene are sequentially expressed in the developing embryo  
A;Reference number: S16317; MUID:91293104; PMID:1676674  
A;Accession: S16318  
A;Molecule type: DNA  
A;Residues: 1-323 <IZP>  
A;Cross-references: UNIPROT:P23813; UNIPARC:UPI0000029F10; EMBL:X60761; NID:G51420; PIDN:G51420

R;Gerard, M.; Duboule, D.; Zakany, J. EMBO J. 12, 3539-3550, 1993  
A;Title: Structure and activity of regulatory elements involved in the activation of the Hox gene cluster

A;Reference number: S40403; MUID:94074553; PMID:7902810  
A;Accession: S40403  
A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-323 <GER>

A;Cross-references: UNIPARC:UPI0000029F10; EMBL:X71422; NID:g937508; PIDN:CAA50553.1; PI  
R;Takada, S.; Cook, M.; Kramlau, R.; McMahon, A.P.  
submitted to the EMBL Data Library, May 1991  
A;Description: Genomic sequence of mouse Hox-4.6.  
A;Reference number: S57443  
A;Accession: S57443  
A;Molecule type: DNA  
A;Residues: 'MNDFCGSAAS', 1-323 <TAK>  
A;Cross-references: UNIPARC:UPI0000026A38; EMBL:X60395; NID:g871427; PIDN:CAA42943.1; PI  
C;Genetics:  
A;Gene: Hoxd-11  
A;Introns: 246/1  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;252-308/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 323;  
Best Local Similarity 92.3%; Pred. No. 18;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 93 AAAAAAAAAAAAAA 105

RESULT 16  
AB1781  
hypothetical protein lin2792 [imported] - Listeria innocua (strain Clip11262)  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AB1781  
R;Glasier, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AB1781  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-328 <GLA>  
A;Cross-references: UNIPROT:Q927J6; UNIPARC:UPI00000CC998; GB:AL592022; PIDN:CAC98018.1;  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin2792

Query Match 88.7%; Score 47; DB 2; Length 328;  
Best Local Similarity 84.6%; Pred. No. 19;  
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 313 AAAAAAAAAAAAAA 325

RESULT 17  
B47236  
zinc-finger protein Pur-1 - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C;Accession: B47236  
R;Kennedy, G.C.; Rutter, W.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992  
A;Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates  
A;Reference number: A47236; MUID:93087555; PMID:1454839  
A;Accession: B47236  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-331 <KEN>  
A;Cross-references: UNIPARC:UPI000017C615  
A;Experimental source: insulinoma cell line T  
A;Note: sequence extracted from NCBI backbone (NCBIP:119832)

Query Match 88.7%; Score 47; DB 2; Length 331;  
Best Local Similarity 92.3%; Pred. No. 19;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 83 AAAAAAAAAAAAAA 95

RESULT 18  
A39065  
homeotic protein EVX2 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 31-Dec-2004  
C;Accession: A39065  
R;D'Esposito, M.; Morelli, F.; Acampora, D.; Migliaccio, E.; Simeone, A.; Boncinelli, E.  
Genomics 10, 43-50, 1991  
A;Title: EVX2, a human homeobox gene homologous to the even-skipped segmentation gene,  
A;Reference number: A39065; MUID:91257849; PMID:1675198  
A;Accession: A39065  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-333 <DAE>  
A;Cross-references: UNIPARC:UPI000017A29D  
C;Genetics:  
A;Gene: GDB:EVX2  
A;Cross-references: GDB:127528; OMIM:142991  
A;Map position: 2q44.3-2q41  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;46-102/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 333;  
Best Local Similarity 92.3%; Pred. No. 19;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 213 AAAAAAAAAAAAAA 225

RESULT 19  
G02409  
protein kinase C-binding protein RACK17 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 05-Nov-1999  
C;Accession: G02409  
R;Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Kikkawa, U.  
submitted to the EMBL Data Library, February 1996  
A;Reference number: H01212  
A;Accession: G02409  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-334 <KUR>  
A;Cross-references: UNIPARC:UPI000016A2D0; EMBL:U48250; NID:g1199656; PIDN:AAC72247.1;

Query Match 88.7%; Score 47; DB 2; Length 334;  
Best Local Similarity 92.3%; Pred. No. 19;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
DB 236 AAAAAAAAAAAAAA 248

RESULT 20  
I48188  
gene Nkx6.1 protein - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Oct-2004  
C;Accession: I48188  
R;Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.  
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994

A;Title: Pancreatic beta cells express a diverse set of homeobox genes.  
A;Reference number: I48185; MUID:95083670; PMID:7991607  
A;Accession: I48188  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-364 <RES>  
A;Cross-references: UNIPROT:Q60554; UNIPARC:UPI000001301FA; EMBL:X81409; NID:g587466; PID  
C;Genetics:  
A;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;237-293/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 364;  
Best Local Similarity 92.3%; Pred. No. 20;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 137 AAAAAAAAAAAAA 149

RESULT 21  
T03875  
probable homeobox protein OSH45, splice form OSH45 [similarity] - rice  
C;Species: Oryza sativa (rice)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Oct-2004  
C;Accession: T03875  
R;Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Mat  
Plant J. 7, 927-938, 1995  
A;Title: Alternative RNA products from a rice homeobox gene.  
A;Reference number: Z15126; MUID:95322999; PMID:7599652  
A;Accession: T03875  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-374 <TAM>  
A;Cross-references: UNIPROT:P93423; UNIPARC:UPI00000AB610; EMBL:D49704; NID:g1805615; PI  
A;Experimental source: cv. Nipponbare  
C;Comment: For alternative splice forms, see PIR:T03874.  
C;Genetics:  
A;Gene: H45  
A;Introns: 174/3; 215/1; 277/3; 321/3; 367/2  
A;Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus  
F;291-352/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 374;  
Best Local Similarity 92.3%; Pred. No. 21;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 64 AAAAAAAAAAAAA 76

RESULT 22  
T03874  
probable homeobox protein OSH45, splice form OSH44 [similarity] - rice  
N;Contains: probable homeobox protein OSH45, splice form OSH42  
C;Species: Oryza sativa (rice)  
C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-Oct-2004  
C;Accession: T03874; T03876  
R;Tamaoki, M.; Tsugawa, H.; Minami, E.; Kayano, T.; Yamamoto, N.; Kano-Murakami, Y.; Mat  
Plant J. 7, 927-938, 1995  
A;Title: Alternative RNA products from a rice homeobox gene.  
A;Reference number: Z15126; MUID:95322999; PMID:7599652  
A;Accession: T03874  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-375 <TAM1>  
A;Cross-references: UNIPROT:P93424; UNIPARC:UPI00000AA5D5; EMBL:D49704; NID:g1805615; PI  
A;Experimental source: cv. Nipponbare; splice form OSH44  
A;Accession: T03876  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA

A;Residues: 197-367 <TAM2>  
A;Cross-references: UNIPARC:UPI000016DEBB; EMBL:D49704; NID:g1805615; PIDN:BAA08554.1; P  
A;Experimental source: cv. Nipponbare; splice form OSH42  
C;Comment: For an alternative splice form, see PIR:T03875.  
C;Genetics:  
A;Gene: H45  
A;Introns: 174/3; 215/1; 277/3; 321/3; 367/2  
A;Keywords: alternative initiators; alternative splicing; DNA binding; homeobox; nucleus  
F;1-375/Product: probable homeobox protein OSH45, splice form OSH44 #status predicted <M  
F;197-367/Product: probable homeobox protein OSH45, splice form OSH42 #status predicted  
F;291-352/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 2; Length 375;  
Best Local Similarity 92.3%; Pred. No. 21;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 64 AAAAAAAAAAAAA 76

RESULT 23  
A44443  
basic helix-loop-helix transcription repressor H - fruit fly (Drosophila virilis)  
C;Species: Drosophila virilis  
C;Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Jul-1998  
C;Accession: A44443  
R;Mainwright, S.M.; Ish-Horowicz, D.  
Mol. Cell. Biol. 12, 2475-2483, 1992  
A;Title: Point mutations in the Drosophila hairy gene demonstrate in vivo requirements f  
A;Reference number: A44443; MUID:92269819; PMID:1588951  
A;Accession: A44443  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-378 <WAI>  
A;Cross-references: UNIPARC:UPI000017BEF1  
A;Note: sequence inconsistent with the nucleotide translation  
A;Note: sequence extracted from NCBI backbone (NCBIN:102964, NCBIP:102966)  
C;Genetics:  
A;Gene: FlyBase:Dvir/h  
A;Cross-references: FlyBase:FBgn0013115

Query Match 88.7%; Score 47; DB 2; Length 378;  
Best Local Similarity 92.3%; Pred. No. 21;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 269 AAAAAAAAAAAAA 281

RESULT 24  
A48423  
engrailed homeodomain-containing protein En-1 - mouse  
N;Alternate names: homeotic protein En-1  
C;Species: Mus musculus (house mouse)  
C;Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C;Accession: A48423; S13009; A26629; A24778  
R;Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.  
Dev. Genet. 13, 345-358, 1992  
A;Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene  
A;Reference number: A48423; MUID:93185339; PMID:1363401  
A;Accession: A48423  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-401 <LOG>  
A;Cross-references: UNIPROT:P09065; UNIPARC:UPI00000299A1  
A;Experimental source: CD-1, embryo  
A;Note: sequence extracted from NCBI backbone (NCBIP:126620)  
R;Holland, P.W.H.; Williams, N.A.  
FEBS Lett. 277, 250-252, 1990  
A;Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution.  
A;Reference number: S13009; MUID:91099509; PMID:1980115

A;Accession: S13009  
A;Status: preliminary  
A;Molecule type: nucleic acid  
A;Residues: 321-380 <HOL>  
A;Cross-references: UNIPARC:UPI000017A2AF  
R;Joyner, A.L.; Martin, G.R.  
Genes Dev. 1, 29-38, 1987  
A;Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrail  
A;Reference number: A91620; MUID:88112776; PMID:2892757  
A;Accession: A26629  
A;Molecule type: DNA; mRNA  
A;Residues: 278-401 <JOY>  
A;Cross-references: UNIPARC:UPI000016CA7C; GB:Y00201; GB:M11987; NID:g49587; PIDN:CAA683  
R;Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.  
Cell 43, 29-37, 1985  
A;Title: Expression during embryogenesis of a mouse gene with sequence homology to the D  
A;Reference number: A24778; MUID:86079501; PMID:2416459  
A;Accession: A24778  
A;Molecule type: DNA  
A;Residues: 311-401 <JO2>  
A;Cross-references: UNIPARC:UPI000017A2B0  
C;Genetics:  
A;Gene: en.1  
A;Map position: 1  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;313-369/Domain: homeobox homology <HOX>

	Query Match	88.7%	Score 47;	DB 2;	Length 401;
	Best Local Similarity	92.3%	Pred. No. 22;		
	Matches 12;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
|||  
Db 207 AAAAAAAAAAAAAA 219

RESULT 25  
A53662  
homeotic protein HB9 - human  
C;Species: Homo sapiens (man)  
C;Date: 08-Jul-1995 #sequence\_revision 03-Aug-1995 #text\_change 31-Dec-2004  
C;Accession: A53662  
R;Harrison, K.A.; Druey, K.M.; Deguchi, Y.; Tuscano, J.M.; Kehrl, J.H.  
J. Biol. Chem. 269, 19968-19975, 1994  
A;Title: A novel human homeobox gene distantly related to proboscipedia is expressed in  
A;Reference number: A53662; MUID:94327547; PMID:7914194  
A;Accession: A53662  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-403 <HAR>  
A;Cross-references: UNIPARC:UPI000017A29E; GB:U07663  
A;Note: the nucleotide sequence and conceptual translation as given are self-consistent  
C;Genetics:  
A;Gene: GDB:HLXB9  
A;Cross-references: GDB:136411; OMIM:142994  
A;Map position: 1q41-1q42.1  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;244-300/Domain: homeobox homology <HOX>

	Query Match	88.7%	Score 47;	DB 2;	Length 403;
	Best Local Similarity	92.3%	Pred. No. 22;		
	Matches 12;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
|||  
Db 121 AAAAAAAAAAAAAA 133

RESULT 26  
T14911  
bZIP DNA-binding protein - parsley  
C;Species: Petroselinum crispum (parsley)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text change 31-Dec-2004

A;Status: preliminary  
A;Molecule type: DNA; mRNA  
A;Residues: 1-475 <DUS>  
A;Cross-references: UNIPROT:P49749; UNIPARC:UPI0000020BC6; GB:S34322; NID:9249137; PTDN:  
A;Note: sequence extracted from NCBI backbone (NCBIN:100478, NCBIN:100480, NCBIN:100482,  
C;Genetics:  
A;Gene: evx-2  
A;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;192-248/Domain: homeobox homology <HOX>  
  
Query Match 88.7%; Score 47; DB 2; Length 475;  
Best Local Similarity 92.3%; Pred. No. 24;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 359 AAAAAAAAAAAAAA 371  
  
RESULT 29  
A47236  
zinc-finger protein Pur-1 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C;Accession: A47236  
R;Kennedy, G.C.; Rutter, W.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992  
A;Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates  
A;Reference number: A47236; MUID:93087555; PMID:1454839  
A;Accession: A47236  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: nucleic acid  
A;Residues: 1-477 <KEN>  
A;Cross-references: UNIPARC:UPI000017C88F  
A;Experimental source: islet cell line beta TC3  
A;Note: sequence extracted from NCBI backbone (NCBIP:119831)  
  
Query Match 88.7%; Score 47; DB 2; Length 477;  
Best Local Similarity 92.3%; Pred. No. 25;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 96 AAAAAAAAAAAAAA 108  
  
RESULT 30  
JC7583  
basic helix-loop-helix protein, DEC2 - human  
C;Species: Homo sapiens (man)  
C;Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004  
C;Accession: JC7583  
R;Fujimoto, K.; Shen, M.; Noshiro, M.; Matsubara, K.; Shingu, S.; Honda, K.; Yoshida, E.  
Biochem. Biophys. Res. Commun. 280, 164-171, 2001  
A;Title: Molecular cloning and characterization of DEC2, a new member of basic helix-loop-helix  
A;Reference number: JC7583; MUID:21092582; PMID:11162494  
A;Accession: JC7583  
A;Molecule type: mRNA  
A;Residues: 1-482 <FUU>  
A;Cross-references: UNIPROT:Q9C0J9; UNIPARC:UPI0000052B52; DBSJ:AB044088  
C;Comment: This protein, a novel member of the DEC subfamily of basic helix-loop-helix F  
C;Genetics:  
A;Gene: dec2  
A;Map position: 12p11.23-p12.1  
C;Keywords: transcription factor  
F;1-173/Region: highly conserved #status predicted  
F;130-173/Domain: Orange #status predicted <ORA>  
F;286-411/Region: alanine and glycine-rich #status predicted  
  
Query Match 88.7%; Score 47; DB 2; Length 482;  
Best Local Similarity 92.3%; Pred. No. 25;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 396 AAAAAAAAAAAAAA 408  
  
RESULT 31  
T09084  
phosphatidylinositol 3-kinase - Chlamydomonas reinhardtii (fragment)  
C;Species: Chlamydomonas reinhardtii  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: T09084  
R;Molendijk, A.J.; Irvine, R.F.  
Plant Mol. Biol. 37, 53-56, 1998  
A;Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinositol  
A;Reference number: Z16411; MUID:98281574; PMID:9620264  
A;Accession: T09084  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-490 <MOL>  
A;Cross-references: UNIPROT:O04270; UNIPARC:UPI000009C6FA; EMBL:U97663; NID:92109290; PI  
A;Experimental source: strain cw-15  
C;Genetics:  
A;Introns: 265/3; 331/3; 370/3; 455/1; 481/3  
  
Query Match 88.7%; Score 47; DB 2; Length 490;  
Best Local Similarity 92.3%; Pred. No. 25;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 115 AAAAAAAAAAAAAA 127  
  
RESULT 32  
A42170  
zinc finger protein MAZ - human (fragment)  
N;Alternate names: MYC-associated zinc finger protein MAZ; zinc finger protein 2F87  
C;Species: Homo sapiens (man)  
C;Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 09-Jul-2004  
C;Accession: A42170; A46153  
R;Pyrc, J.J.; Moberg, K.H.; Hall, D.J.  
Biochemistry 31, 4102-4110, 1992  
A;Title: Isolation of a novel cDNA encoding a zinc-finger protein that binds to two sites  
A;Reference number: A42170; MUID:92232709; PMID:1567856  
A;Accession: A42170  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-494 <PYR>  
A;Cross-references: UNIPROT:P56270; UNIPARC:UPI000017C427; GB:J05371  
A;Note: it is uncertain whether Met-18 is the initiator or whether translation is initiated  
R;Bossone, S.A.; Asselin, C.; Patel, A.J.; Marcu, K.B.  
Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456, 1992  
A;Title: MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating tra  
A;Reference number: A46153; MUID:92366479; PMID:1502157  
A;Accession: A46153  
A;Molecule type: mRNA  
A;Residues: 18-417, 'L', 419-494 <BOS>  
A;Cross-references: UNIPARC:UPI000012ECF8; GB:M94046  
A;Experimental source: HeLa cells  
A;Note: sequence extracted from NCBI backbone (NCBIN:110666, NCBIP:110667)  
C;Keywords: DNA binding; zinc finger  
F;113-125/Region: alanine-rich  
F;174-183/Region: zinc finger  
F;207-230/Region: zinc finger  
F;296-318/Region: zinc finger  
F;324-346/Region: zinc finger  
F;354-368/Region: zinc finger  
F;373-405/Region: zinc finger  
F;409-430/Region: zinc finger  
F;452-468/Region: alanine-rich  
  
Query Match 88.7%; Score 47; DB 2; Length 494;  
Best Local Similarity 92.3%; Pred. No. 25;



```
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 113 AAAAAAAAAAAAAA 125

RESULT 33
S31223
transcription factor Brn-1 - mouse
N:Alternate names: class III POU domain protein brain-1
C:Species: Mus musculus (house mouse)
C:Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S31223
R:Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992
A:Title: Structure and evolution of four POU domain genes expressed in mouse brain.
A:Reference number: S31223; MUID:92228768; PMID:1565620
A:Accession: S31223
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-495 <HAR>
A:Cross-references: UNIPROT:P31361; UNIPARC:UPI0000029AAF; EMBL:M88299; NID:G200444; PID
C:Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:101-112/Region: alanine-rich
F:162-180/Region: histidine/proline-rich
F:186-201/Region: alanine-rich
F:236-247/Region: glycine-rich
F:267-291/Region: histidine/proline-rich
F:316-383/Domain: POU domain homology <POU>
F:402-458/Domain: homeobox homology <HOX>

Query Match 88.7%; Score 47; DB 1; Length 495;
Best Local Similarity 92.3%; Pred. No. 25;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 186 AAAAAAAAAAAAAA 198

RESULT 34
JC5076
myc-associated zinc-finger protein - human
N:Alternate names: MAZ protein
C:Species: Homo sapiens (man)
C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 05-Nov-1999
C:Accession: JC5076
R:Tsutsui, H.; Sakatsume, O.; Itakura, K.; Yokoyama, K.K.
Biochem. Biophys. Res. Commun. 226, 801-809, 1996
A:Title: Members of the MAZ family: A novel cDNA clone for MAZ from human pancreatic isl
A:Reference number: JC5076; MUID:96428591; PMID:8831693
A:Accession: JC5076
A:Molecule type: mRNA
A:Residues: 1-497 <TSU>
A:Cross-references: UNIPARC:UPI0000163B39; DBJ:D85131; NID:g1752741; PIDN:BAAL2728.1; F
A:Experimental source: pancreatic islet
C:Comment: This protein plays a role in the control of transcriptional initiation of gen
and between the introns of the mouse gene for immunoglobulin M-D.
C:Keywords: phosphoprotein; zinc finger
F:146,204,480/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status pre
F:349/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 88.7%; Score 47; DB 2; Length 497;
Best Local Similarity 92.3%; Pred. No. 25;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 122 AAAAAAAAAAAAAA 134

RESULT 35
A48233
polyomavirus enhancer-binding protein 2 alpha chain type 1 - mouse
N:Alternate names: PEA2 alpha chain type 1; PEA2 alpha chain type 2; PEBP2 alpha chain
C:Species: Mus musculus (house mouse)
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 09-Jul-2004
C:Accession: A48233; B48233
R:Ogawa, E.; Maruyama, M.; Kagoshima, H.; Inuzuka, M.; Lu, J.; Satake, M.; Shigesada, K.
Proc. Natl. Acad. Sci. U.S.A. 90, 6859-6863, 1993
A:Title: PEBP2/PEA2 represents a family of transcription factors homologous to the prod
A:Reference number: A48233; MUID:93342088; PMID:8341710
A:Accession: A48233
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-513 <OGA>
A:Cross-references: UNIPROT:Q08775; UNIPARC:UPI000002B2F8; GB:D14636; NID:g391766; PIDN
A:Accession: B48233
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-304,'L',306 <OG2>
A:Cross-references: UNIPARC:UPI000002B2F9; GB:D14637; NID:g391768; PIDN:BAA03486.1; PID
C:Genetics:
A:Gene: PEBP2alpha
C:Superfamily: transcription factor CBF alpha 2
C:Keywords: alternative splicing; DNA binding; T-cell; transcription factor; transcript

Query Match 88.7%; Score 47; DB 2; Length 513;
Best Local Similarity 92.3%; Pred. No. 26;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 64 AAAAAAAAAAAAAA 76

RESULT 36
A55929
zinc finger protein noca - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 09-Jul-2004
C:Accession: A55929
R:Cheah, P.Y.; Meng, Y.B.; Yang, X.; Kimbrell, D.; Ashburner, M.; Chia, W.
Mol. Cell. Biol. 14, 1487-1499, 1994
A:Title: The Drosophila 1(2)35Ba/noca gene encodes a putative Zn finger protein involve
A:Reference number: A55929; MUID:94119100; PMID:8289824
A:Accession: A55929
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-537 <CHE>
A:Cross-references: UNIPROT:Q24423; UNIPARC:UPI000007B356; GB:L14009; NID:g431289; PID:
C:Genetics:
A:Gene: FlyBase:noc
A:Cross-references: FlyBase:FBgn0005771

Query Match 88.7%; Score 47; DB 2; Length 537;
Best Local Similarity 84.6%; Pred. No. 27;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 232 SSMAAAAAAAAAAA 244

RESULT 37
WJFFEN
homeotic protein engrailed - fruit fly (Drosophila melanogaster)
N:Alternate names: specific body pattern development protein
C:Species: Drosophila melanogaster
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 09-Jul-2004
C:Accession: A90862; A93354; A03321; A25682; S03667
R:Poole, S.J.; Kauvar, L.M.; Drees, B.; Kornberg, T.
Cell 40, 37-43, 1985
```

A;Title: The engrailed locus of Drosophila: structural analysis of an embryonic transcript  
A;Reference number: A90862; MUID:85099327; PMID:3917855  
A;Accession: A90862  
A;Molecule type: mRNA  
A;Residues: 1-552 <PO>  
A;Cross-references: UNIPROT:P02836; UNIPARC:UPI000012CA13; GB:M10017; NID:g157363; PIDN:  
R;Jose, A.; McGinnis, W.J.; Gehring, W.J.  
Nature 313, 284-289, 1985  
A;Title: Isolation of a homeo box-containing gene from the engrailed region of Drosophila  
A;Reference number: A93354; MUID:90114393; PMID:2481829  
A;Accession: A93354  
A;Molecule type: DNA  
A;Residues: 447-485, 'E', 487-519, 'WH' <FJO>  
A;Cross-references: UNIPARC:UPI00001745BF; GB:X01765; GB:K03059; NID:98084; PIDN:CA2590  
A;Note: the translation of the nucleotide sequence from Fig. 5 is shown, not the translation in Fig. 6  
R;Kassir, J.A.; Poole, S.J.; Wright, D.K.; O'Farrell, P.H.  
EMBO J. 5, 3583-3589, 1986  
A;Title: Sequence conservation in the protein coding and intron regions of the engrailed  
A;Reference number: A91059; MUID:87161768; PMID:2881781  
A;Contents: annotation; intron locations and sequences  
R;Gay, N.J.; Poole, S.J.; Kornberg, T.B.  
Nucleic Acids Res. 15, 6637-6647, 1988  
A;Title: The Drosophila engrailed protein is phosphorylated by a serine-specific protein  
A;Reference number: S03667; MUID:88289425; PMID:2899884  
A;Contents: annotation; potential phosphorylation sites; homeobox domain  
A;Comment: This protein specifies the body segmentation pattern.  
C;Genetics:  
A;Gene: en  
A;Cross-references: FlyBase:FBgn0000577  
A;Map position: 2R,62.0 (48A1-4)  
A;Introns: 438/1; 470/3  
C;Superfamily: engrailed homeotic protein; homeobox homology  
C;Keywords: DNA binding; embryo; homeobox; nucleus; segmentation; transcription regulation  
F;26-53/Region: glutamine-rich  
F;68-87/Region: alanine-rich  
F;232-240/Region: alanine-rich  
F;320-368/Region: serine-rich  
F;455-511/Domain: homeobox homology <HOX>  
  
Query Match 88.7%; Score 47; DB 1; Length 552;  
Best Local Similarity 92.3%; Pred. No. 27;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 68 AAAAAAAAAAAAAA 80  
  
RESULT 38  
T02610  
probable YME1 ATP-dependant proteinase [imported] - Arabidopsis thaliana  
N;Alternate names: YTA11 protein homolog T19L18.5  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 31-Dec-2004  
C;Accession: T02610; H84656  
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,  
submitted to the EMBL Data Library, August 1998  
A;Description: Arabidopsis thaliana chromosome II BAC T19L18 genomic sequence.  
A;Reference number: Z14681  
A;Accession: T02610  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-627 <ROU>  
A;Cross-references: UNIPROT:O80983; UNIPARC:UPI000017885F; EMBL:AC004747; NID:g3413696;  
A;Experimental source: cultivar Columbia  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umavam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: H84656

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-627 <STO>  
A;Cross-references: UNIPARC:UPI000017885F; GB:AE002093; NID:g3413700; PIDN:AAC31223.1; G  
C;Genetics:  
A;Gene: T19L18.5; At2g26140  
A;Map position: 2  
A;Introns: 40/1; 119/1; 160/3; 319/3  
C;Superfamily: FtsH/SEC18/CDC48-type ATP-binding domain homology  
  
Query Match 88.7%; Score 47; DB 2; Length 627;  
Best Local Similarity 92.3%; Pred. No. 30;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 598 AAAAAAAAAAAAAA 610  
  
RESULT 39  
S39356  
transcription factor btd - fruit fly (Drosophila sp.)  
C;Species: Drosophila sp.  
C;Date: 18-Feb-1994 #sequence\_revision 01-Dec-1995 #text\_change 07-May-1999  
C;Accession: S39356  
R;Wimmer, E.A.; Jaekle, H.; Pfeifle, C.; Cohen, S.M.  
Nature 366, 690-694, 1993  
A;Title: A Drosophila homologue of human Spl is a head-specific segmentation gene.  
A;Reference number: S39356; MUID:94081952; PMID:8259212  
A;Accession: S39356  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-644 <WIN>  
A;Cross-references: UNIPARC:UPI0000124C17; EMBL:Z29361; NID:g441283; PID:g441284  
C;Genetics:  
A;Gene: FlyBase:btd  
A;Cross-references: FlyBase:FBgn0000233  
A;Introns: 245/2  
  
Query Match 88.7%; Score 47; DB 2; Length 644;  
Best Local Similarity 92.3%; Pred. No. 31;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
||| ||||| |||||  
Db 205 AAAAAAAAAAAAAA 217  
  
RESULT 40  
S40382  
box A-binding factor - fruit fly (Drosophila melanogaster)  
N;Alternate names: ABF; transcription factor dGATAB  
C;Species: Drosophila melanogaster  
C;Date: 13-Jan-1995 #sequence\_revision 06-Sep-1996 #text\_change 09-Jul-2004  
C;Accession: S40382  
R;Abel, T.; Michelson, A.M.; Maniatis, T.  
Development 119, 623-633, 1993  
A;Title: A Drosophila GATA family member that binds to Adh regulatory sequences is expressed  
A;Reference number: S40382; MUID:94244465; PMID:8187633  
A;Accession: S40382  
A;Molecule type: mRNA  
A;Residues: 1-779 <ABE>  
A;Cross-references: UNIPROT:P52172; UNIPARC:UPI0000135F5F; EMBL:X76217; NID:g441491; PID  
C;Comment: This transcriptional activator is the earliest known marker of the developing  
C;Genetics:  
A;Gene: FlyBase:srp  
A;Cross-references: FlyBase:FBgn0003507  
C;Superfamily: box A-binding factor; GATA-type zinc finger homology  
C;Keywords: DNA binding; nucleus; transcription factor; transcription regulation; zinc  
F;316-369/Domain: GATA-type zinc finger homology <GZF>  
F;319-343/Region: zinc finger GATA motif  
  
Query Match 88.7%; Score 47; DB 1; Length 779;

Best Local Similarity 92.3%; Pred. No. 35;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAA 13  
| | | | | | | | | |  
Db 602 AAAAAAAAAAAA 614

Search completed: September 9, 2006, 23:01:49  
Job time : 16.8101 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:38:21 ; Search time 119.139 Seconds  
(without alignments)  
100.934 Million cell updates/sec

Title: US-10-617-568-3

Perfect score: 53

Sequence: 1 AAAAAAAAAAAAA 13

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

UniProt\_7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	484	1	ZBTB8_MOUSE
2	53	100.0	484	2	Q3US18_MOUSE
3	49	92.5	167	2	Q6Z8T9_ORYZA
4	49	92.5	361	2	Q4PE18_USTMA
5	49	92.5	368	2	Q7SBK2_NEUCR
6	49	92.5	377	1	H5F7_ARATH
7	49	92.5	378	2	Q9NXD6_HUMAN
8	49	92.5	395	2	Q80US0_MOUSE
9	49	92.5	395	2	Q8BKCI_MOUSE
10	49	92.5	395	2	Q3TSG4_MOUSE
11	49	92.5	395	2	Q8BK9_MOUSE
12	49	92.5	428	1	FOXK2_MOUSE
13	49	92.5	432	2	Q5VYV0_HUMAN
14	49	92.5	435	2	Q6LGI1_XENTR
15	49	92.5	458	2	Q8P6C2_HUMAN
16	49	92.5	506	2	Q3S2W8_ACHDO
17	49	92.5	512	1	ZBTB8_HUMAN
18	49	92.5	521	2	Q7PPS0_ANOGA
19	49	92.5	529	2	Q4P290_USTMA
20	49	92.5	578	1	SOB_DROME
21	49	92.5	640	2	Q9VTW5_DROME
22	49	92.5	708	2	Q9NVR0_HUMAN
23	49	92.5	708	2	Q8BR71_MOUSE
24	49	92.5	709	2	Q8CE33_MOUSE
25	49	92.5	805	2	Q4SPH0_TETNG
26	49	92.5	815	1	PYGO_DROME
27	49	92.5	828	2	Q3UG42_MOUSE
28	49	92.5	892	2	O70254_MOUSE
29	49	92.5	897	2	Q9VE88_DROME
30	49	92.5	914	2	Q69Z11_MOUSE
31	49	92.5	935	2	Q3UYA4_MOUSE

## ALIGNMENTS

32	49	92.5	946	2	Q9VI93_DROME
33	49	92.5	964	1	LBXCO_MOUSE
34	49	92.5	964	1	LBXCO_RAT
35	49	92.5	965	1	LBXCO_HUMAN
36	49	92.5	1065	2	Q4SPK5_TETNG
37	49	92.5	1181	2	Q8IR69_DROME
38	49	92.5	1199	2	Q59E49_DROME
39	49	92.5	1203	2	Q819J7_DROME
40	49	92.5	1248	2	Q819J6_DROME
41	49	92.5	1255	2	Q819J6_DROME
42	49	92.5	1298	2	Q5BIAT_DROME
43	49	92.5	1533	1	PUM_DROME
44	49	92.5	1324	2	Q2KML0_RAT
45	49	92.5	1325	2	Q2KML1_RAT
46	49	92.5	1927	2	Q2KMK9_RAT
47	49	92.5	1959	2	Q2KMK7_RAT
48	49	92.5	1976	2	Q4QHY5_LEIMA
49	49	92.5	2000	1	CHD3_HUMAN
50	49	92.5	2038	1	FSH_DROME
51	49	92.5	2055	2	Q5NCG1_MOUSE
52	49	92.5	3012	2	O97205_LEIMA
53	49	92.5	3941	2	Q623D7_CABER
54	48	90.6	293	2	O17222_CAREL
55	47	88.7	52	2	Q56ZW8_ARATH
56	47	88.7	71	2	Q8NI53_HUMAN
57	47	88.7	112	2	Q3UUS5_MOUSE
58	47	88.7	117	2	Q3V062_MOUSE
59	47	88.7	130	2	O9TUC9_CANFA
60	47	88.7	131	2	O9TST8_FELCA
61	47	88.7	132	2	Q4TBK8_TETNG
62	47	88.7	143	2	Q2YHB4_CHICK
63	47	88.7	145	2	Q2YX3_ANNPU
64	47	88.7	149	2	Q7Q8K9_ANOGA
65	47	88.7	150	2	O9Y4M1_HUMAN
66	47	88.7	153	2	Q765Q0_RAT
67	47	88.7	157	2	Q62GX4_BURMA
68	47	88.7	158	2	Q9MBF7_LILLO
69	47	88.7	164	2	Q7Q9M3_ANOGA
70	47	88.7	164	2	Q8BNH5_MOUSE
71	47	88.7	167	2	Q765P9_RAT
72	47	88.7	171	2	Q8H8G6_ORYZA
73	47	88.7	175	2	Q7XRMI_ORYZA
74	47	88.7	191	2	Q96N13_HUMAN
75	47	88.7	199	2	Q4RK77_TETNG
76	47	88.7	203	2	Q5SC14_HUMAN
77	47	88.7	213	2	O9TUC8_MONDO
78	47	88.7	218	2	Q5JG20_HUMAN
79	47	88.7	220	2	Q61PH7_HUMAN
80	47	88.7	223	2	Q5ISR0_MACFA
81	47	88.7	227	2	Q7Q706_ANOGA
82	47	88.7	228	2	Q5K4L2_9MYRI
83	47	88.7	233	1	ASCL1_RAT
84	47	88.7	233	2	Q7XTV6_ORYZA
85	47	88.7	236	1	ASCL1_HUMAN
86	47	88.7	236	2	Q6QDA4_HUMAN
87	47	88.7	236	2	Q9BX46_HUMAN
88	47	88.7	244	2	Q6CNB7_KLULA
89	47	88.7	247	2	Q9DDF6_PETMA
90	47	88.7	251	2	Q9VHD5_DROME
91	47	88.7	253	2	Q8WZ46_HUMAN
92	47	88.7	260	2	Q95T44_DROME
93	47	88.7	266	2	Q963L0_DROME
94	47	88.7	269	2	Q4JHB6_HUMAN
95	47	88.7	272	2	Q753B0_ASHGO
96	47	88.7	276	1	SOX21_HUMAN
97	47	88.7	276	1	SOX21_MOUSE
98	47	88.7	276	2	Q5TBS1_HUMAN
99	47	88.7	280	1	SOX21_CHICK
100	47	88.7	284	2	Q6EEZ3_9ARAC

Q9VI93	drosophila
Q8BX46	mus musculus
P84551	rattus norv
P84550	homo sapien
Q4SPK5	tetradon n
Q8IR69	drosophila
Q59E49	drosophila
Q819J7	drosophila
Q9VI93	drosophila
Q819J6	drosophila
Q5BIAT	drosophila
P28222	drosophila
Q2KML0	rattus norv
Q2KML1	rattus norv
Q2KMK9	rattus norv
Q2KMK7	rattus norv
Q4QHY5	leishmania
Q12873	homo sapien
P13709	drosophila
P13709	drosophila
Q5NCG1	mus musculus
O97205	leishmania
Q623D7	caenorhabdi
O17222	caenorhabdi
Q56ZW8	arabidopsis
Q8NI53	homo sapien
Q3UUS5	mus musculus
Q3V062	mus musculus
O9TUC9	canis famil
O9TST8	felis silve
Q4TBK8	tetradon n
Q2YHB4	gallus gall
Q2YX3	annitella pu
Q7Q8K9	anopheles g
O9Y4M1	homo sapien
Q765Q0	rattus norv
Q62GX4	burkholderi
Q9MBF7	lillium long
Q7Q9M3	anopheles g
Q8BNH5	mus musculus
Q765P9	rattus norv
Q8H8G6	oryza sativ
Q7XRMI	oryza sativ
Q96N13	homo sapien
Q4RK77	tetradon n
O5SC14	homo sapien
O9TUC8	monodelphis
Q5JG20	homo sapien
Q61PH7	homo sapien
Q5ISR0	macaca fasc
Q7Q706	anopheles g
Q5K4L2	glomeris ma
P13359	rattus norv
Q7XTV6	oryza sativ
P50553	homo sapien
Q6QDA4	homo sapien
Q9BX46	homo sapien
Q6CNB7	kluyveromyc
Q9DDF6	petromycon
Q9VHD5	drosophila
Q8WZ46	homo sapien
Q95T44	drosophila
Q963L0	drosophila
Q4JHB6	homo sapien
Q753B0	ashbya goss
O9Y651	homo sapien
Q811W0	mus musculus
O97755	mus musculus
Q9W755	gallus gall
Q6EEZ3	euprosthen

Best Local Similarity 100.0%; Pred. No. 42;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
| | | | | | | | | |  
Db 136 AAAAAAAAAAAAA 148

## RESULT 2

Q3US18 MOUSE  
ID Q3US18 MOUSE PRELIMINARY; PRT; 484 AA.  
AC Q3US18  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 21-FEB-2006, entry version 7.  
DE 16 days embryo head cDNA, RIKEN full-length enriched library, full  
DE clone: C130068L16 product: zinc finger and BTB domain containing 8, full  
DE insert sequence (0 day neonate cerebellum cDNA, RIKEN full-length  
DE enriched library, clone: C230068H07 product: zinc finger and BTB domain  
DE RIKEN full-length enriched library, clone: 3222401B20 product: zinc  
DE finger and BTB domain containing 8, full insert sequence).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=927253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning."  
RL Methods Enzymol. 303:19-44(1999).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C., Sheng Y.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Sultana R., Takenaka Y., Taki K.,  
RA Shering S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tamoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Ikeda J., Inamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,

## RESULT 1

ZBTB8 MOUSE  
ID ZBTB8 MOUSE STANDARD; PRT; 484 AA.  
AC O8CII0;  
DT 26-APR-2004, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAR-2003, sequence version 1.  
DT 07-MAR-2006, entry version 22.  
DE Zinc finger and BTB domain-containing protein 8.  
GN Name=Zbtb8;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=FVB/N;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Besak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalys D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RL and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: May be involved in transcriptional regulation.  
CC -!- SUBCELLULAR LOCATION: Nucleus (potential).  
CC -!- SIMILARITY: Contains 1 BTB (POZ) domain.  
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC023839.1; -; mRNA.  
DR Ensembl; ENSMUSG0000048485; Mus musculus.  
DR MGI; MGI:2387181; Zbtb8.  
DR InterPro; IPR000210; BTB.  
DR InterPro; IPR013069; BTB\_POZ.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zF-C2H2; 2.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; Znf\_C2H2; 2.  
DR PROSITE; PS50097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 2.  
KW DNA-binding; Metal-binding; Nuclear protein; Repeat; Transcription;  
KW Transcription regulation; Zinc; Zinc-finger;  
FT CHAIN 1 484  
FT Zinc finger and BTB domain-containing  
FT protein 8.  
FT /FTID=PRO\_0000047722.  
FT BTB.  
FT ZN\_FING 331 353 C2H2-type 1.  
FT ZN\_FING 359 382 C2H2-type 2.  
FT COMPBIAS 132 149 Ala-rich.  
SQ SEQUENCE 484 AA; 53387 MW; B4FF082555B0CA1A CRC64;  
Query Match 100.0%; Score 53; DB 1; Length 484;

RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563 (2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX PubMed:16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
(RIKEN Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome";  
RL Science 309:1564-1566 (2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=22354683; PubMed=12468951; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sadelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verdaro R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yaunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Mateu Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boirelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690 (2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630 (2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=20330913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Fuyukawa Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771 (2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK140914; BAE24519.1; -; mRNA.  
DR EMBL; AK163930; BAE37541.1; -; mRNA.  
DR EMBL; AK132152; BAE20999.1; -; mRNA.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005515; F:protein binding; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR InterPro; IPR000210; BTF.  
DR InterPro; IPR013069; BTF POZ.  
DR InterPro; IPR007087; Znf C2H2.  
DR Pfam; PF00651; BTF; 1.  
DR Pfam; PF00096; zf-C2H2; 2.  
DR ProDom; PD000003; Znf C2H2; 1.  
DR SMART; SM00225; BTF; 1.  
DR SMART; SM00355; Znf C2H2; 2.  
DR PROSITE; PS00097; BTF; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 2.  
DR Metal-binding; Nuclear protein; Repeat; Zinc; Zinc-finger.  
SQ SEQUENCE 484 AA; 53414 MW; ACA81D5402AE7F4 CRC64;  
  
Query Match 100.0%; Score 53; DB 2; Length 484;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAAAAA 13  
Db 136 AAAAAAAAAAAAAA 148

```

RESULT 3
ID Q628T9_ORYSA PRELIMINARY; PRT; 167 AA.
AC Q628T9;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 14.
DE Zinc finger protein family-like.
GN Names=P0686H1.17;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzoideae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- DOMAIN: The RING-type zinc finger domain is essential for
CC ubiquitin ligase activity. It coordinates an additional third zinc
CC ion (by similarity).
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AP004762; BAD10011.1; -; Genomic_DNA.
DR Gramine; Q628T9;
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO: GO:0046872; F:metal ion binding; IEA.
DR GO: GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0016567; P:protein ubiquitination; IEA.
DR InterPro: IPR011016; RINGv.
DR InterPro: IPR001841; Znf_RING.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR SMART: SM00744; RINGv; 1.
DR PROSITE: PS00089; ZF_RING_2; 1.
DR Metal-binding; Nuclear protein; Zinc; Zinc-finger.
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.
SQ SEQUENCE 167 AA; 17664 MW; 4729ACE00A901AE9 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 167;
Best Local Similarity 92.3%; Pred. No. 57;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
DB 56 AAAAAAAAAAAAAA 68
[[:::|||||]]

RESULT 4
ID Q4PEI8_USTMA PRELIMINARY; PRT; 361 AA.
AC Q4PEI8;
DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Hypothetical protein.
GN ORFNames=UM01475.1;
OS Ustilago maydis 521.
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=237631;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=521;
RA Birren B.W., Nussbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H.M., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
```

```

Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S.E., Camarata J., Campo K., Chang J., Cheshatsang Y.,
RA Citroen M., Collymore A., Considine T., Cook A., Cooke P., Corum B.,
RA Cuomo C., David R., Dawce T., Degray S., Dodge S., Dooley K.,
RA Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,
RA Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J.E., Garin G., Gnerre S.,
RA Gnrke A., Goyette A., Graham J., Grandbois E., Gvaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseles M., Karlsson E.,
RA Kells C., Kieu A., Kishner P., Kodira C., Kulbokas E., Labutti K.,
RA Lanna D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-Toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.-J., Mabbitt R., Maru K., Matthews C., Mauceil E.,
RA Manning J., Marabelli R., McGhee T., Meldrim J., Meneus L.,
RA McCarthy M., McDonough S., McGhee T., Mikkelsen T., Mlenga V., Moru K.,
RA Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neves C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C.B., Nizzari M., Norbu C.,
RA Norbu N., O'Donnell P., Okoawo O., O'Leary S., Omotosho B., Pignani B.,
RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-Thomann N., Stavropoulos S.,
RA Stetson K., Stone S., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA Towey S., Tsamila T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V.S., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee B., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.S.;
RT "The genome sequence of Ustilago maydis";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AACP0100053; EAK82242.1; -; Genomic_DNA.
DR Hypothetical protein.
SQ SEQUENCE 361 AA; 39615 MW; 038F774408202041 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 361;
Best Local Similarity 92.3%; Pred. No. 1e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
DB 133 AAAAAAAAAAAAAA 145
[[:::|||||]]

RESULT 5
ID Q7SBK2_NEUCR PRELIMINARY; PRT; 368 AA.
AC Q7SBK2;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE Predicted protein.
GN ORFNames=NCU08548.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=74-OR23-1A / FGSC 987;
RA MEDLINE=22598136; PubMed=12712197; DOI=10.1038/nature01554;
RX
```

RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,  
RA Jaffe D., Fitchugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,  
RA Elkins T., Engels P., Wang S., Nielsen C.B., Butler J., Endrizzi M.,  
RA Qui D., Iankiev P., Bell-Pedersen D., Nelson M.A.,  
RA Werner-Washburne M., Selitrennikoff C.P., Kinsey J.A., Braun E.L.,  
RA Zelter A., Schulte U., Koth G.O., Jedd G., Mewes H.-W., Staben C.,  
RA Marcotte E., Greenberg D., Roy A., Foley K., Naylor J.,  
RA Stange-Thomann N., Barrett R., Gnerre S., Kamal M., Kamyselis M.,  
RA Maucelli E., Bielke C., Rudd S., Frishman D., Krystofova S.,  
RA Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S., Cogoni C.,  
RA Macino G., Catcheside D.E.A., Li W., Pratt R.J., Osmari S.A.,  
RA Desouza C.P.C., Glass N.L., Orbach M.J., Berglund J.A., Voelker R.,  
RA Yarden O., Plamann M., Seiler S., Dunlap J.C., Radford A., Aramayo R.,  
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbola D.J., Freitag M.,  
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaume C., Birren B.W.;  
RA "The genome sequence of the filamentous fungus *Neurospora crassa*.";  
RT Nature 422:859-868(2003).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AABX01000166; EAA33794.1; -; Genomic DNA.  
SQ SEQUENCE 368 AA; 35691 MW; 7AB0F69255959DD2 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 368;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AMAAAAA 13  
Db 255 AMAAAAA 266

RESULT 6  
ID HSF7 ARATH STANDARD; PRT; 377 AA.  
AC Q970D3; Q9SCW3;  
DT 01-DEC-2000, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAY-2000, sequence version 1.  
DT 07-MAR-2006, entry protein 7.  
DE Heat shock factor protein 7 (HSF 7) (Heat shock transcription factor  
DE 7) (HSF7).  
DN Name=HSF7; OrderedLocusNames=At4g11660; ORFNames=TF5C23.90;  
GN Arabidopsis thaliana (Mouse-ear cress).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
OX rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20083488; PubMed=10617198; DOI=10.1038/47134;  
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,  
RA Pohl T., Duysterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,  
RA Weichselbarger M., de Simone V., Obermaier B., Mache R., Mueller M.,  
RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,  
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
RA Vos P., Hobeisel J., Zimmermann W., Wedler H., Ridley P.,  
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,  
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
RA Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,  
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
RA Moeljan P., Klein lankhorst R., Rose M., Hauf J., Kotter P.,  
RA Berneriser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,  
RA De Keyser A., Buysshaert C., Gielens J., Villarroel R., De Clercq R.,  
RA Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,  
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,  
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,

RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,  
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Bauner D., Herzl A.,  
RA Neumann S., Argicou A., Vitale D., Liguori R., Piravandi E.,  
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,  
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
RA Chefod F., Cooke R., Berger C., Monfort A., Casacuberta E.,  
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,  
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
RA Farnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,  
RA Sekhom T., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,  
RA Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,  
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,  
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,  
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,  
RA Chen E., Marra M.A., Martienssen R., McCombie W.R.;  
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 402:769-777(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=cv. Columbia;  
RX MEDLINE=22954850; PubMed=14593172; DOI=10.1126/science.1088305;  
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
RA Arakawa T., Barh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
RA Hayaishizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
RA Yuan S., Shinohara K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Empirical analysis of transcriptional activity in the Arabidopsis  
RT genome.";  
RL Science 302:842-846(2003).  
RN [3]  
RP NUCLEOTIDE SEQUENCE OF 50-377.  
RC STRAIN=cv. Columbia; TISSUE=Green siliques;  
RA Schoeffl F., Fraendl R.,  
RT "De-repression of heat shock protein synthesis in transgenic plants.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: DNA-binding protein that specifically binds heat shock  
CC promoter elements (HSE) and activates transcription (By  
CC similarity).  
CC -!- SUBUNIT: Homotrimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nucleus (By similarity).  
CC -!- PTM: Exhibits temperature-dependent phosphorylation (By  
CC similarity).  
CC -!- SIMILARITY: Belongs to the HSP family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AL049500; CAB3937.1; -; Genomic DNA.  
DR EMBL; AL161532; CAB78209.1; -; Genomic DNA.  
DR EMBL; AY093206; AAM13205.1; -; mRNA.  
DR EMBL; BT008471; AAP37830.1; -; mRNA.  
DR EMBL; AJ251868; CAB63803.1; -; mRNA.  
DR F04213; T04213.  
DR HSP; P22813; 1HKT.  
DR GenomeReviews; CT486007\_GR; AT4G11660.



DR GeneFarm; 4072; 412.  
DR TAIR; At4g11660; -.  
DR InterPro; IPR000232; HSF DNA bd.  
DR InterPro; IPR002341; HSF ETS DNA bd.  
DR InterPro; IPR011991; Wing hlx DNA bd.  
DR Pfam; PF00447; HSF DNA-bind; 1.  
DR PRINTS; PR00056; HSF DNA-bind; 1.  
DR PRODOM; PD001788; HSF DNA\_bind; 1.  
DR SMART; SM00415; HSF; 1.  
DR PROSITE; PS00434; HSF DOMAIN; 1.  
KW Activator; DNA-binding; Heat shock; Nuclear protein; Phosphorylation;  
KW Transcription; Transcription regulation.  
FT CHAIN 1 377  
FT Heat shock factor protein 7.  
FT FTID=PRO\_0000124588.  
FT  
FT DNA BIND 57 151  
FT By similarity.  
FT COMPBIAS 14 21  
FT Poly-Gly.  
FT COMPBIAS 32 50  
FT Poly-Gly.  
FT COMPBIAS 155 171  
FT Poly-Ala.  
FT COMPBIAS 202 207  
FT Poly-Ala.  
FT CONFLICT 50 50  
FT G -> R (in Ref. 3).  
SQ SEQUENCE 377 AA; 39705 MW; FD1D5E595B7FC584 CRC64;  
  
Query Match 92.5%; Score 49; DB 1; Length 377;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 AAAAAAAAAAAAA 13  
Db 155 AAAAAAAAAAAAA 166  
  
RESULT 7  
Q9NXD6\_HUMAN PRELIMINARY; PRT; 378 AA.  
AC Q9NXD6;  
DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2000, sequence version 1.  
DT 07-FEB-2006, entry version 18.  
DE Hypothetical protein FLJ20308.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,  
RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isonaga S., Sugano S.;  
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK000315; BAA91078.1; -; mRNA.  
DR Ensembl; ENSG0000091542; Homo sapiens.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR InterPro; IPR005123; 2OG-FeII\_Oxy.  
DR Pfam; PF03171; 2OG-FeII\_Oxy; 1.  
KW Iron; Oxidoreductase.  
SQ SEQUENCE 378 AA; 42533 MW; 1E7E5C741ACF6BA6 CRC64;  
  
Query Match 92.5%; Score 49; DB 2; Length 378;  
Best Local Similarity 92.3%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAAAA 13  
Db 20 AAAAAAAAAAAAA 32  
  
RESULT 8

Q80US0\_MOUSE PRELIMINARY; PRT; 395 AA.  
AC Q80US0;  
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2003, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE Hypothetical protein Ofoxd.  
GN Name=Ofoxd;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Mouse;  
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.A., McEwan P.J., McKernan K.J., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smailus D.E.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.;  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6; TISSUE=Mouse;  
RG NIH MGC Project;  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC052076; AAH52076.1; -; mRNA.  
DR Ensembl; ENSMUSG0000042650; Mus musculus.  
DR MGI; MGI:2144489; AW050020.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR InterPro; IPR005123; 2OG-FeII\_Oxy.  
DR Pfam; PF03171; 2OG-FeII\_Oxy; 1.  
KW Hypothetical protein; Iron; Oxidoreductase.  
SQ SEQUENCE 395 AA; 44410 MW; 4CE25B4FBSB04DEB CRC64;  
  
Query Match 92.5%; Score 49; DB 2; Length 395;  
Best Local Similarity 92.3%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAAAA 13  
Db 36 AAAAAAAAAAAAA 48  
  
RESULT 9  
Q8BK1\_MOUSE PRELIMINARY; PRT; 395 AA.  
AC Q8BK1;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 22.  
DE 0 day neonate eyeball cDNA, RIKEN full-length enriched library.

DE clone:EL30207K11 product:hypothetical Alanine-rich region/Type I  
DE antifreeze protein/20G-Fe(II) oxygenase superfamily containing  
DE protein, full insert sequence (2 cells egg cDNA, RIKEN full-length  
DE enriched library, clone:B020011005 product:Hypothetical alanine-rich  
DE region/type I antifreeze protein/20G-Fe, full insert sequence) (Novel  
DE protein)  
DE Name=AW050020; ORFNames=RP23-331I21.1-001;  
GN Mus musculus (Mouse).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN (2)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama K., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Bremner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Aweiler R., Aturaliya R.N., Bailey T.L.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakamura H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugiuira K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN (3)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN (4)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;

RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Oato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragan T.A., Fletcher C.F., Forrest A., Fraser K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai J., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa A.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
RN (5)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=25781660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Stauber J., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
RN (6)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630(2000).  
RN (7)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Egg, and Eyeball;  
RX MEDLINE=205330913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada K.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,

```
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer.";
RN Genome Res. 10:1757-1771(2000).
[8]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Eyeball;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hirakata T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[9]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Egg;
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
[10]
RP NUCLEOTIDE SEQUENCE.
RA Tracey A.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR ENBL; AK053695; BAC35478.1; -; mRNA.
DR ENBL; AK163294; BAE37281.1; -; mRNA.
DR ENBL; AL596386; CA135332.1; -; Genomic DNA.
DR ENBL; ENSMUSG0000042650; Mus musculus.
DR MGI; MGI:2144489; AW050020.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR InterPro; IPR005123; 2OG-Fell_Oase.
DR Pfam; PF03171; 2OG-Fell_Oxy; I
DR Hypothetical protein; Ifon; Oxidoreductase.
SQ SEQUENCE 395 AA; 44411 MW; 2822BF4BFFFE7EF CRC64;

Query Match 92.5%; Score 49; DB 2; Length 395;
Best Local Similarity 92.3%; Pred. No. 1.1e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAAAAA 13
Db 36 AAAAAAAAAAAAAA 48

RESULT 10
Q3TSG4 MOUSE
AC Q3TSG4 PRELIMINARY; PRT; 395 AA.
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, entry version 1.
DE In vitro fertilized eggs cDNA, RIKEN full-length enriched library,
DE clone:7420402E17 product:Hypothetical alanine-rich region/type I
DE antifreeze protein/2OG-Fe, full insert sequence.
GN Name=AW050020;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
```

```
OK NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RN Methods Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Iwama A., Ishikawa T.,
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuoka H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugizawa K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX PubMed=16141073; DOI=10.1126/science.1112009;
RG RIKEN Genome Exploration Research Group, and Genome Science Group
RG (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
[4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nozaki A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazar K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
```

RA Konagawa A., Kurochkin I. V., Lee Y., Lenhard B., Lyons P. A.,  
RA Maglott D.R., Maltais K., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Nunata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempole C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Glissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsumoto Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690 (2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630 (2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771 (2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=In vitro fertilized eggs;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,

RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK162072; BAE36711.1; -; mRNA.  
DR MGI; MGI:2144489; AW050020.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR InterPro; IPR005123; 2OG-Ferri\_Oase.  
DR Pfam; PF03171; 2OG-Ferri\_Oxy; I.  
DR Hypothetical protein.  
SQ SEQUENCE 395 AA; 44494 MW; 1F3B93E0E71AF27D CRC64;  
  
Query Match 92.5%; Score 49; DB 2; Length 395;  
Best Local Similarity 92.3%; Pred.No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAAAAA 13  
DB 36 AAAAAAAAAAAAAA 48  
||:|||||  
||:|||||  
  
RESULT 11  
Q8BK9\_MOUSE PRELIMINARY; PRT; 395 AA.  
ID Q8BK9\_MOUSE  
AC Q8BK9;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-MAR-2006, entry version 24.  
DE 0 day neonate eyeball cDNA, RIKEN full-length enriched library.  
DE clone:EI3021021 product:hypothetical Alanine-rich region/Type I  
DE antifreeze protein/2OG-Fe(II) oxygenase superfamily containing  
DE protein, full insert sequence.  
DE Name:AW050020;  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning";  
RL Methods Enzymol. 303:19-44 (1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler R., Attalaya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Georger C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Geich-Hemminger P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,

RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salberg S.L., Sandelin A., Schneider C., Sheng Y.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptome landscape of the mammalian genome.";  
RL Science 309:1559-1563 (2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome";  
RL Science 309:1564-1566 (2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573 (2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690 (2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630 (2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771 (2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Eyeball;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC EMBL; AK053700; BAC35481.1; -; mRNA.  
DR Ensembl; ENSMUSG0000042650; Mus musculus.  
DR MGI; MGI:2144489; AW050020.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
KW Hypothetical protein.  
SQ SEQUENCE 395 AA; 44415 MW; 6D2C26B1A99EDA8 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 395;  
Best Local Similarity 92.3%; Pred. No. 1.1e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
DB 36 AAAAAAAAAAAAAA 48

RESULT 12  
FOX2\_MOUSE  
ID FOXB2\_MOUSE STANDARD; PRT; 428 AA.

AC Q64733; 01-NOV-1997, integrated into UniProtKB/Swiss-Prot.  
 DT 01-NOV-1997, sequence version 1.  
 DT 07-FEB-2006, entry version 31.  
 DE Forkhead box protein B2 (Transcription factor FKX-4).  
 GN Name=Foxb2; Synonyms=Fkh4;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=97014266; PubMed=8861101; DOI=10.1016/0925-4773(96)00507-2;  
 RA Kaestner K.H., Schuetz G., Monaghan A.P.;  
 RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains  
 in the central nervous system.";  
 RL Mech. Dev. 55:221-230(1996).  
 CC [2]  
 CC NUCLEOTIDE SEQUENCE OF 4-114.  
 CC STRAIN=129;  
 CC MEDLINE=93361500; PubMed=7689224;  
 RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H., Monaghan A.P.,  
 RA Schuetz G.;  
 RT "Six members of the mouse forkhead gene family are developmentally  
 regulated.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).  
 CC -!- SUBCELLULAR LOCATION: Nucleus.  
 CC -!- DEVELOPMENTAL STAGE: Expressed during embryogenesis.  
 CC -!- SIMILARITY: Contains 1 fork-head DNA-binding domain.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: X92591; CAA63335.1; -; mRNA.  
 DR EMBL: X71942; CAA50744.1; -; Genomic\_DNA.  
 DR PIR: D47746; D47746.  
 DR HSSP: Q63245; 2HDC.  
 DR TRANSFAC: T02442; -.  
 DR Ensembl: ENSMUSG0000056829; Mus musculus.  
 DR MGI: MGI:1347468; Foxb2.  
 DR InterPro: IPR001766; TF\_Fork head.  
 DR InterPro: IPR011991; Wing\_hlx\_DNA\_bd.  
 DR Pfam: PF00250; Fork head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR ProDom: PD000425; TF\_Fork\_head; 1.  
 DR SMART: SM00339; FH; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; 1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.  
 FT CHAIN 1 428 Forkhead box protein B2.  
 FT /FTID=PRO\_0000091805.  
 FT Fork-head.  
 FT DNA BIND 12 103  
 FT COMPBIAS 139 153 Poly-His.  
 FT COMPBIAS 156 162 Poly-His.  
 FT COMPBIAS 163 172 Poly-Pro.  
 FT COMPBIAS 217 231 Poly-Ala.  
 FT COMPBIAS 249 258 Poly-Ala.  
 FT COMPBIAS 321 330 Poly-Ala.  
 FT COMPBIAS 396 399 Poly-Ala.  
 SQ SEQUENCE 428 AA; 45170 MW; DB8A8EFD1E94AB10 CRC64;  
 Query Match 92.5%; Score 49; DB 1; Length 428;  
 Best Local Similarity 92.3%; Pred. No. 1.2e+02;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 Db 218 AAAAAAAAAAAAAA 230  
 RESULT 13

Q5YVVO HUMAN  
 ID Q5YVVO\_HUMAN PRELIMINARY; PRT; 432 AA.  
 AC Q5YVVO;  
 DT 07-DEC-2004, integrated into UniProtKB/TrEMBL.  
 DT 07-DEC-2004, sequence version 1.  
 DE OTTHUMP0000021510.  
 DE OTTHUMP0000021510.  
 GN ORFNames=RP11-159H20.4-001;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Kimberley A.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL: AL353637; CAH70683.1; -; Genomic\_DNA.  
 DR GO: GO:0005634; C:nucleus; IEA.  
 DR GO: GO:0003700; F:transcription factor activity; IEA.  
 DR GO: GO:0006350; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO: GO:0006350; P:transcription; IEA.  
 DR InterPro: IPR001766; TF\_Fork head.  
 DR InterPro: IPR011991; Wing\_hlx\_DNA\_bd.  
 DR Pfam: PF00250; Fork head; 1.  
 DR PRINTS: PR00053; FORKHEAD.  
 DR ProDom: PD000425; TF\_Fork\_head; 1.  
 DR SMART: SM00339; FH; 1.  
 DR PROSITE: PS00657; FORK\_HEAD\_1; 1.  
 DR PROSITE: PS00658; FORK\_HEAD\_2; UNKNOWN\_1.  
 DR PROSITE: PS50039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.  
 SQ SEQUENCE 432 AA; 45581 MW; 9A832ACDD9A765EF CRC64;  
 Query Match 92.5%; Score 49; DB 2; Length 432;  
 Best Local Similarity 92.3%; Pred. No. 1.2e+02;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 Db 221 AAAAAAAAAAAAAA 233  
 RESULT 14  
 Q6GLG1 XENTR  
 ID Q6GLG1\_XENTR PRELIMINARY; PRT; 435 AA.  
 AC Q6GLG1;  
 DT 19-JUL-2004, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2004, sequence version 1.  
 DE 07-FEB-2006, entry version 16.  
 DE SWI/SNF related, matrix associated, actin dependent regulator of  
 chromatin, subfamily e, member 1.  
 GN Name=smarcel1-prov;  
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodidae; Xenopus; Silurana.  
 NCBI\_TaxID=8364;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Embryo;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]

RNA NUCLEOTIDE SEQUENCE.

RP TISSUE=Brain;

RG NIH MGC Project;

RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

CC EMBL: BC062339; AAH62339.1; -; mRNA.

DR Ensembl: ENSG0000091542; Homo sapiens.

DR GO: GO:0005506; F:iron ion binding; IEA.

DR GO: GO:0016491; F:oxidoreductase activity; IEA.

DR InterPro: IPR005123; 2OG-FelI\_Oase.

DR Pfam: PF03171; 2OG-FelI\_Oxy; 1.

DR KW Hypothetical protein; Iron; Oxidoreductase.

DR SEQUENCE 458 AA; 51409 MW; 88ACCD9992B5AEB CRC64;

QY

Query Match 92.5%; Score 49; DB 2; Length 458;

Best Local Similarity 92.3%; Pred. No. 1.3e+02;

Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db

QY 1 AAAAAAAAAAAAAA 13

36 AAAAAAAAAAAAAA 48

||:|||||

RESULT 16

Q3S2W8\_ACHDO PRELIMINARY; PRT; 506 AA.

AC Q3S2W8;

DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.

DT 11-OCT-2005, sequence version 1.

DT 21-FEB-2006, entry version 7.

DE Broad21 isoform.

OS Acheta domesticus (House cricket).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Orthopteroidea; Orthoptera; Ensifera; Gryllidae; Gryllinae;

OC Acheta.

OX NCBI\_TaxID=6997;

OP [1]

RNA NUCLEOTIDE SEQUENCE.

RP Erezylmaz D.F., Riddiford L.M., Truman J.W.;

RA "An ancestral role for a metamorphosis-determining factor in a direct-

RT developing insect.";

RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.

CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC -----

CC EMBL: DQ176003; ABA02190.1; -; mRNA.

DR GO: GO:0005634; C:nucleus; IEA.

DR GO: GO:0003676; F:nucleic acid binding; IEA.

DR GO: GO:0005515; F:protein binding; IEA.

DR GO: GO:0008270; F:zinc ion binding; IEA.

DR InterPro: IPR000210; BTB.

DR InterPro: IPR013069; BTB\_POZ.

DR InterPro: IPR07087; Znf\_C2H2.

DR Pfam: PF00651; BTB; 1.

DR Pfam: PF00096; zf-C2H2; 2.

DR SMART: SM00225; BTB; 1.

DR SMART: SM00355; Znf\_C2H2; 2.

DR PROSITE: PS50097; BTB; 1.

DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 2.

DR PROSITE: PS50157; ZINC\_FINGER\_C2H2\_2; 2.

KW Metal-Binding; Nuclear protein; Zinc; Zinc-finger.



SQ SEQUENCE 506 AA; 54175 MW; F4FCCA9BB6415EC0 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 506;  
 Best Local Similarity 92.3%; Pred. No. 1.4e+02;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAMAAARAAAAA 13  
 ||:|||||  
 Db 218 AAVAAAAA 230

RESULT 17

ZBTB8\_HUMAN  
 ID ZBTB8\_HUMAN STANDARD; PRT; 512 AA.  
 AC Q8NAP8; Q5VXR5;  
 DT 26-APR-2004, integrated into UniProtKB/Swiss-Prot.  
 DT 01-OCT-2002, sequence version 1.  
 DT 07-MAR-2006, entry version 25.  
 DE Zinc finger and BTB domain-containing protein 8.  
 GN Name=ZBTB8;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
 OC Homo.  
 OX NCBI\_TaxID=9606;  
 RN [(1)]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
 RC TISSUE=Brain;  
 RX PubMed=14702039; DOI=10.1038/ng1285;  
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
 RA Sekine M., Ohbayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
 RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
 RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
 RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
 RA Abe K., Kamiyama K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
 RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y., Ishida S.,  
 RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
 RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
 RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
 RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
 RA Yoshihara Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
 RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Satoh N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 cDNAs.";  
 RL Nat. Genet. 36:40-45(2004).  
 RN [(2)]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RG Human chromosome 1 international sequencing consortium;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: May be involved in transcriptional regulation.  
 CC -!- SUBCELLULAR LOCATION: Nucleus (Potential).  
 CC -!- SIMILARITY: Contains 1 BTB (POZ) domain.  
 CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AK092326; BAC03863.1; -; mRNA.

DR EMBL; AL356986; CAH73365.1; -; Genomic DNA.  
 DR EMBL; AL033529; CAH73365.1; JOINED; Genomic DNA.  
 DR EMBL; AL033529; CAI20022.1; -; Genomic DNA.  
 DR EMBL; AL356986; CAI20022.1; JOINED; Genomic DNA.  
 DR Ensembl; ENSG00000185718; Homo sapiens.  
 DR HGNC; HGNC:24172; ZBTB8.  
 DR InterPro; IPR000210; BTB.  
 DR InterPro; IPR013069; BTB\_POZ.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00651; BTB; 1.  
 DR Pfam; PF00096; zf-C2H2; 2.  
 DR ProDom; PD000003; Znf\_C2H2; 1.  
 DR SMART; SM00225; BTB; 1.  
 DR SMART; SM00355; Znf\_C2H2; 2.  
 DR PROSITE; PS00097; BTB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
 KW DNA-binding; Metal-binding; Nuclear protein; Repeat; Transcription;  
 FT Transcription regulation; Zinc; Zinc-finger.  
 FT CHAIN 1 512 Zinc finger and BTB domain-containing  
 FT protein 8.  
 FT /FTID=PRO\_0000047721.  
 FT DOMAIN 24 92 BTB  
 FT ZN\_FING 341 363 C2H2-type 1.  
 FT ZN\_FING 369 392 C2H2-type 2.  
 FT COMPBIAS 132 153 Poly-Ala.  
 SQ SEQUENCE 512 AA; 55562 MW; BB970D200B60CC9F CRC64;

Query Match 92.5%; Score 49; DB 1; Length 512;  
 Best Local Similarity 92.3%; Pred. No. 1.4e+02;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAMAAARAAAAA 13  
 ||:|||||  
 Db 136 AAVAAAAA 148

RESULT 18

Q7PPS0\_ANOGA  
 ID Q7PPS0\_ANOGA PRELIMINARY; PRT; 521 AA.  
 AC Q7PPS0;  
 DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.  
 DT 07-DEC-2004, sequence version 2.  
 DT 07-FEB-2006, entry version 16.  
 DE ENSANGP0000016235  
 GN ORFNames=ENSANGG000000013746;  
 OS Anopheles gambiae str. PEST.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicidae;  
 OC Anophelinae; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [(1)]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=PEST;  
 RG The Anopheles gambiae Sequence Committee;  
 RT "Anopheles gambiae re-annotation.";  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [(2)]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=PEST;  
 RG The Anopheles gambiae Sequence Committee;  
 RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AA001008933; EAA09944.3; -; Genomic DNA.  
 DR GO; GO:0003676; P:nucleic acid binding; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR InterPro; IPR012677; a\_b\_plait\_nuc\_bd.



```
DR InterPro; IPR006532; PolyU bd.
DR InterPro; IPR00504; RNPI_RNA_bd.
DR InterPro; IPR003954; RRM_1.
DR PANTHER; PTHR10432:SF141; PolyU_half_pint; 2.
DR Pfam; PF00076; RRM_1; 3.
DR SMART; SM00360; RRM; 3.
DR SMART; SM00361; RRM_1; 1.
DR TIGRFAMS; TIGR01645; half_pint; 1.
DR PROSITE; PS0102; RRM; 3.
SQ SEQUENCE 521 AA; 55760 MW; 68A0C7DEA9FE24DD CRC64;

Query Match 92.5%; Score 49; DB 2; Length 521;
Best Local Similarity 92.3%; Pred. No. 1.4e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
   ||:|||||
Db 339 AAAAAAAAAAAAAA 351

RESULT 19
Q4P290 USTMA PRELIMINARY; PRT; 529 AA.
AC Q4P290; 2005, integrated into UniProtKB/TrEMBL.
DT 19-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 4.
DE Hypothetical protein.
GN ORFNames=UM05773.1;
OS Ustilago maydis 521.
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=237631;
RN [1]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=521;
RA Birren B.W., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Alt-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Atchachi H.M., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S.E., Camarata J., Campo K., Chang J., Cheshatsang Y.,
RA Citroen M., Collymore A., Considine T., Cooke P., Corum B.,
RA Cuomo C., David R., Dawoe T., Degray S., Dodge S., Dooley K.,
RA Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,
RA Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgeraid M., Foley K., Gage D., Galagan J.E., Gearin G., Gnerre S.,
RA Girke A., Goyette A., Graham J., Grandbois E., Gyalteen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseles M., Karlsson E.,
RA Kells C., Kieu A., Kiser P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Landblad-Toh K., Liu X., Lokytasang T., Lokytasang Y., Lucien O.,
RA Lui A., Ma L.-J., Mabbitt R., MacDonald J., MacLean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., McGhee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihalev A., Minova T., Mikkelsen T., Menga V., Moru K.,
RA Moraes J., Mulrain L., Munson R., Naylor J., News C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C.B., Nizzari M., Norbu C.,
RA O'Neill K., O'Donnell P., Okawa O., O'Leary S., Omotosho B.,
RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piquani B.,
RA Purcell S., Rachupka T., Ramaamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson R., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schnupbach R., Seaman C., Settippalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-Thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Valle D., Vassiliev H., Wangchuk T.,
RA Venkataraman V.S., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whitaker C., Wilkinson J., Wu Y., Wyman D., Yadvav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
```

```
RA Zimmer A., Zody M., Lander E.S.;
RT "The genome sequence of Ustilago maydis.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AACP01000212; EAX86012.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000679; Znf_GATA.
DR Pfam; PF00320; GATA; 1.
DR SMART; SM00401; Znf_GATA; 1.
DR PROSITE; PS01114; GATA_ZN_FINGER_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 529 AA; 53820 MW; DSABC2C8EA96B06D CRC64;

Query Match 92.5%; Score 49; DB 2; Length 529;
Best Local Similarity 92.3%; Pred. No. 1.4e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
   ||:|||||
Db 76 AAAAAAAAAAAAAA 88

RESULT 20
SOB_DROME
ID SOB_DROME STANDARD; PRT; 578 AA.
AC Q9VQS7; Q24571;
DT 16-AUG-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-MAY-2000, sequence version 1.
DT 07-MAR-2006, entry version 30.
DE Protein sister of odd and bowel.
GN Name=sob; ORFNames=CG3242;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
NUCLEOTIDE SEQUENCE [MRNA], AND TISSUE SPECIFICITY.
RC STRAIN=Canton-S; TISSUE=Embryo;
EX MEDLINE=97032935; PubMed=8878683;
RA Hart M.C., Wang L., Coulter D.E.;
RT "Comparison of the structure and expression of odd-skipped and two
RT related genes that encode a new family of zinc finger proteins in
RT Drosophila.";
RL Genetics 144:171-182 (1996);
RN [2]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
```

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.C., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.C., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D., Weinstock G.M., Weissensbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*";  
RT Science 287:2185-2195(2000).  
RN [3]  
RN GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bertram B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [4]  
RN NCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=Berkley; TISSUE=Embryo;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
RA George R.A., Gonzalez M., Guarin H., Kronmiller B., Li P.W., Liao G.,  
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J.M., Paragas V., Park S.,  
RA Patel S., Phouanavong S., Wan K.H., Yu C., Lewis S.E., Rubin G.M.,  
RA Celniker S.E.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RN POSSIBLE FUNCTION, AND TISSUE SPECIFICITY.  
RP PubMed=14597202; DOI=10.1016/j.ydbio.2003.07.011;  
RA Hao I., Green R.B., Dunaevsky O., Lengyel J.A., Rauskolb C.;  
RT "The odd-skipped family of zinc finger genes promotes *Drosophila* leg  
RL segmentation.";  
RL Dev. Biol. 263:282-295(2003).  
RN [6]  
RN TISSUE SPECIFICITY.  
RP MEDLINE=22930851; PubMed=14568103; DOI=10.1016/j.mod.2003.08.001;  
RA Johansen K.A., Green R.B., Iwaki D.D., Hernandez J.B., Lengyel J.A.;  
RT "The Dm-Bowl-Lin relief-of-repression hierarchy controls fore- and  
RT hindgut patterning and morphogenesis";  
RL Mech. Dev. 120:1139-1151(2003).  
CC -!- FUNCTION: Pair-rule protein that determines both the size and  
CC polarity of even-numbered as well as odd-numbered parasegments  
CC during embryogenesis. DNA-binding transcription factor that acts  
CC primarily as a transcriptional repressor but can also function as  
CC a transcriptional activator, depending on the stage of development  
CC and spatial restrictions (by similarity). May function redundantly  
CC with odd and drm in leg joint formation during the larval stages,  
CC acting downstream of Notch activation.  
CC -!- SUBCELLULAR LOCATION: Nucleus (Probable).  
CC -!- TISSUE SPECIFICITY: Has two temporally distinct modes of  
CC expression during early embryogenesis; expressed in seven stripes  
CC at the blastoderm stage. Also expressed in a non-periodic domain  
CC at the anterior of the embryo. During gastrulation, the seven  
CC primary stripes are supplemented by seven secondary stripes that

appear in alternate segments. This results in the labelling of  
each of the 14 segments in the extended germ band. Expression is  
relatively weak at the blastoderm stage, gaining in intensity at  
gastrulation. Expressed in the invaginating stomodaeum and  
proctodaeum of the embryonic gut. By stage 13, expressed in the  
region that will form the proventriculus and in a wide ring at the  
most posterior portion of the midgut. Expression continues in the  
gut through the remainder of embryogenesis. Expressed in the  
proximal Malpighian tubules, brain and pharyngeal muscles during  
late embryogenesis. Expressed weakly in a segmentally repeated  
pattern in the leg disk at the distal edge of each presumptive leg  
segment except in tarsal segments 1 to 4.  
CC -!- SIMILARITY: Contains 5 C2H2-type zinc fingers.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; U62004; AAC47282.1; -; mRNA.  
CC EMBL; AE003579; AAF51087.1; -; Genomic DNA.  
CC EMBL; BT003205; AAO24960.1; -; mRNA.  
CC PIR; S72227; S72227.  
CC HSP; P07248; 2ADR.  
CC  
CC Ensembl; CG3242; Drosophila melanogaster.  
CC Flybase; FBgn0004892; sob.  
CC GO; GO:0005634; C:nucleus; ISS.  
CC GO; GO:0003677; P:DNA binding; ISS.  
CC GO; GO:0016563; P:transcriptional activator activity; ISS.  
CC GO; GO:0016564; P:transcriptional repressor activity; ISS.  
CC GO; GO:0007350; P:blastoderm segmentation; ISS.  
CC GO; GO:0016348; P:leg joint morphogenesis (sensu Endopterygota); IMP.  
CC GO; GO:0000122; P:negative regulation of transcription from R.; ISS.  
CC GO; GO:0007366; P:periodic partitioning by pair rule gene; ISS.  
CC GO; GO:0045944; P:positive regulation of transcription from R.; ISS.  
CC InterPro; IPR007087; Znf\_C2H2.  
CC Pfam; PF00096; zf-C2H2; 5.  
CC ProDom; PD000003; Znf\_C2H2; 2.  
CC SMART; SM00355; Znf\_C2H2; 5.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 5.  
CC Activator; Complete proteome; Developmental protein; DNA-binding;  
CC Metal-binding; Nuclear protein; Pair-rule protein; Repeat; Repressor;  
CC Transcription; Transcription regulation; Zinc; Zinc-finger.  
CC CHAIN 1 578  
CC /FTID=PRO\_0000046928.  
CC  
CC ZN\_FING 395 417  
CC ZN\_FING 423 445  
CC ZN\_FING 451 473  
CC ZN\_FING 479 501  
CC ZN\_FING 507 529  
CC ZN\_FING 537 559  
CC COMPBIAS 9 303  
CC COMPBIAS 182 256  
CC CONFLICT 176 176  
CC CONFLICT 188 189  
CC CONFLICT 546 546  
CC SEQUENCE 578 AA; 58455 NW; 0F600954CFA7D8D0 CRC64;  
  
CC Query Match 92.5%; Score 49; DB 1; Length 578;  
CC Best Local Similarity 92.3%; Pred. No. 1.5e+02;  
CC Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
CC QY 1 AAAAAAAAAAAAAA 13  
CC ||:|||||  
CC Db 274 AAAAAAAAAAAAAA 286  
  
CC RESULT 21  
CC Q9VTW5 DROME  
CC ID Q9VTW5 DROME PRELIMINARY; PRT; 640 AA.  
CC AC Q9VTW5;  
CC DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
CC DT 01-OCT-2002, sequence version 2.  
CC DT 21-FEB-2006, entry version 39.  
CC DE CG32105-PB (RE70810p).

GN ORFNames-CG32105, Dmel CG32105;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OX Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bock J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.S., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Swirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Swirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Swirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RX FlyBase;  
RA Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Berkley;  
RA Stapleton M., Carlson J., Chavez C., Frise E., George R., Pacleb J.,  
RA Park S., Wan K., Yu C., Celniker S.;  
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: Contains 2 LIM zinc-binding domains.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
Distributed under the Creative Commons Attribution-NoDerivs License  
EMBL; AE003541; AAF49930.2; -; Genomic\_DNA.  
EMBL; BT003467; AAO39470.1; -; mRNA.  
HSSP; P50480; 1BW5.  
INACT; Q9VTW5; -  
DR FlyBase; FBgn0052105; CG32105.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR012287; Homeodomain-rel.  
DR InterPro; IPR007107; LIM homeo.  
DR InterPro; IPR001781; LIM Zn bd.  
DR InterPro; IPR001965; Znf\_PHD.  
DR Pfam; PF00046; Homeobox; 1.  
DR Pfam; PF00412; LIM; 2.  
DR PRINTS; PR00024; HOMEBOX.  
DR ProDom; PD000010; Homeobox; 1.  
DR ProDom; PD000094; LIM; 2.  
DR SMART; SM00389; HOX; 1.  
DR SMART; SM00132; LIM; 2.  
DR PROSITE; PS00027; HOMEBOX\_1; 1.  
DR PROSITE; PS00071; HOMEBOX\_2; 1.  
DR PROSITE; PS00473; LIM DOMAIN 1; 2.  
DR PROSITE; PS00473; LIM DOMAIN 2; 2.  
DR PROSITE; PS01359; ZF\_PHD 1; UNKNOWN 1.  
KW DNA-binding; Developmental protein; Homeobox; LIM domain;  
KW Metal-binding; Nuclear protein; Transcription;  
KW Transcription regulation; Zinc.  
SQ SEQUENCE 640 AA; 67760 MW; ALA10F826018C98E CRC64;  
Query Match 92.5%; Score 49; DB 2; Length 640;  
Best Local Similarity 92.3%; Pred. No. 1.6e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAAA 13  
DB 189 AAAAAAAAAAAAAA 201  
RESULT 22  
Q9NVR0\_HUMAN PRELIMINARY; PRT; 708 AA.  
ID Q9NVR0\_HUMAN  
AC Q9NVR0;

DT 01-OCT-2000, integrated into UniProtKB/TrEMBL.  
DT 01-OCT-2000, sequence version 1.  
DT 07-MAR-2006, entry version 27.  
DE Hypothetical protein FLJ10572 (Kelch-like 11).  
GN Name=KLHL11;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=14702039; DOI=10.1038/ngl285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Ingaki H., Ikema Y., Okamoto S.,  
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Tuzushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RN Nat. Genet. 36:40-45(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Testis;  
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC TISSUE=Testis;  
RX NIH MGC Project;  
RN Submitted (JUL-2002) to the EMBL/GenBank/DBSJ databases.

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
DR EMBL; AK001434; BAA91689.1; -; mRNA.  
DR EMBL; BC034470; AAH34470.1; -; mRNA.  
DR Ensembl; ENSG00000178502; Homo sapiens.  
DR HGNC; HGNC:19008; KLHL11.  
DR GO; GO:0005515; F:protein binding; IEA.  
DR InterPro; IPR011705; BACK.  
DR InterPro; IPR000210; BTB.  
DR InterPro; IPR013069; BTB\_POZ.  
DR InterPro; IPR006652; Kelch\_rep.  
DR Pfam; PF07707; BACK; 1.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF01344; Kelch; 1.  
DR SMART; SM00225; BTB; 1.  
DR PROSITE; PS50097; BTB; 1.  
SQ SEQUENCE 708 AA; 80148 MW; 38733CB875172E12 CRC64;  
  
Query Match 92.5%; Score 49; DB 2; Length 708;  
Best Local Similarity 92.3%; Pred. No. 1.8e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAIAAAAAA 13  
DB 4 AAAAAAIAAAAAA 16  
||:|||||  
||:|||||  
  
RESULT 23  
QB8R71\_MOUSE PRELIMINARY; PRT; 708 AA.  
ID QB8R71\_MOUSE  
AC QB8R71\_MOUSE  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 21.  
DE Adult male corpora quadrigena cDNA, RIKEN full-length enriched  
DE library, clone:B230201M16 product:Mi-2 autoantigen 240 kDa protein  
DE homolog (Fragment).  
GN Name=Chd3;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigena;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigena;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Bremner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Walming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Cioma M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,

RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Hottelgu-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Mori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Suglira K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlstedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,  
RA Tegami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX PubMed=16141073; DOI=10.1126/science.112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
(Rgenome Network Core Team) and the PANTOM Consortium;  
RT "Antisense transcription in the Mammalian Transcriptome.";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Lee Y., Lenhard B., Lyons P.A.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Negashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Pavasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes.";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuami T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RX STRAIN=C57BL/6J; TISSUE=Corpora quadrigemina;  
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kasukawa T.,  
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kouda M., Koya S.,  
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
RA Tagawa A., Takahashi F., Takaku-Akaiura S., Takeda Y., Tanaka T.,  
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
-----  
CC EMBL; AK045449; BAC32375.1; -; mRNA.  
CC HSSP; Q14839; 1NM2.  
CC SMR; Q8BR71; 464-519.  
CC Ensembl; ENSMUSG0000018474; Mus musculus.  
CC MGI; MGI:1344395; Chd3.  
CC GO; GO:0016585; C:chromatin remodeling complex; IDA.  
CC GO; GO:0016581; C:NURD complex; IDA.  
CC GO; GO:0005515; F:protein binding; IPI.  
CC GO; GO:0006333; P:chromatin assembly or disassembly; IDA.  
CC InterPro; IPR012958; CHD N.  
CC InterPro; IPR000953; Chromo.  
CC InterPro; IPR001965; Znf\_PHD.  
CC Pfam; PF08073; CHDNT; 1.  
CC Pfam; PF00385; Chromo; 1.  
CC Pfam; PF06628; PHD; 2.  
CC SMART; SM00298; CHROMO; 2.  
CC SMART; SM00249; PHD; 2.

DR PROSITE, PS00598; CHROMO\_1; UNKNOWN\_1.  
 DR PROSITE, PS50013; CHROMO\_2; 2.  
 DR PROSITE, PS50016; ZF\_PHD\_2; 2.  
 KW Nuclear protein.  
 FT NON TER 708  
 SQ SEQUENCE 708 AA; 80411 MW; A9CE144CE3310CBB CRC64;  
 Query Match 92.5%; Score 49; DB 2; Length 708;  
 Best Local Similarity 92.3%; Pred. No. 1.8e+02;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 Db 218 AAAAAAAAAAAAAA 230  
 RESULT 24  
 Q8CE33\_MOUSE PRELIMINARY; PRT; 709 AA.  
 AC Q8CE33;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 07-FEB-2006, entry version 10.  
 DE 10 days neonate skin cDNA, RIKEN full-length enriched library,  
 DE clone:473491u07 product:hypothetical BT5/POZ domain|Kelch repeat|Type  
 DE 1 antifreeze|Alanine-rich region containing protein, full insert  
 DE sequence.  
 GN Name=Klhl11;  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=16141072; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aldinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Kitano M., Knapin A., Knoch M., Kawasawa Y., Kelso J., Kitamura H.,  
 RA Kikano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Fabriz S., Mulder N., Nakano N., Nakachi H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki N., Orlandi V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling S., Stupka E., Sugitara K., Sultana R., Takeda Y., Taki K.,  
 RA Tannoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yananishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,

RA Grimond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Pleasy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Fraser K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang L., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";   
RL Nature 409:685-690(2001).   
RN [6]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;   
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,   
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;   
RT "Normalization and subtraction of cap-trapper-selected cDNAs to   
RT prepare full-length cDNA libraries for rapid discovery of new genes.";   
RL Genome Res. 10:1617-1630(2000).   
RN [7]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

RP NUCLEOTIDE SEQUENCE.   
RC STRAIN=C57BL/6J; TISSUE=Skin;   
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;   
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,   
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,   
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,   
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,   
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,   
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,   
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;   
RT "RIKEN integrated sequence analysis (RISA) system-384-format   
RT sequencing pipeline with 384 multipillar sequencer.";   
RL Genome Res. 10:1757-1771(2000).   
RN [8]

DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.   
DT 19-JUL-2005, sequence version 1.   
DT 07-MAR-2006, entry version 8.   
DE Chromosome 16 SCAF14537, whole genome shotgun sequence. (Fragment).

GN ORFNames=GSTENG0014837001;   
OS Tetraodon nigroviridis (Green puffer).   
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;   
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;   
OC Tetraodontidae; Tetraodontidae; Tetraodon.   
OX NCBI\_TaxID=99883;   
RN [1]

RP NUCLEOTIDE SEQUENCE.   
RX PubMed=15496914; DOI=10.1038/nature03025;   
RA Jallouf O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,   
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,   
RA Nicaud C., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,   
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,   
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,   
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,   
RA Craud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,   
RA Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,   
RA Kellis M., Volff J.-N., Guigo R., Zody M.C., Mesirov J.,   
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,   
RA Laudet V., Schachter V., Quetier P., Saurin W., Scarpelli C.,   
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;   
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals   
RT the early vertebrate proto-karyotype.";   
RL Nature 431:946-957(2004).   
RN [2]

RP NUCLEOTIDE SEQUENCE.   
RG Genoscope, Whitehead Institute Centre for Genome Research;   
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.   
CC !- CAUTION: The sequence shown here is derived from an   
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is   
CC preliminary data.   
CC !- FUNCTION: Binds calmodulin in a calcium dependent manner. May   
CC function as scaffolding or signaling protein (By similarity).   
CC !- SUBCELLULAR LOCATION: Cytoplasmic and membrane-bound (By   
CC similarity).   
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>   
CC Distributed under the Creative Commons Attribution-NoDerivs License

CC EMBL; CAE01014537; CAF97462.1; -; Genomic DNA.   
DR InterPro; IPR001680; WD40.   
DR Pfam; PF00400; WD40; 6.   
DR PRINTS; PR00320; GPROTEINBRPT.   
DR ProDom; PD000018; WD40; 3.   
DR SMART; SM00320; WD40; 5.   
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.   
DR PROSITE; PS00082; WD\_REPEATS\_2; 4.   
DR PROSITE; PS0294; WD\_REPEATS\_REGION; 1.   
KW Calmodulin-binding; WD repeat.   
FT NON\_TER 1   
SQ SEQUENCE 805 AA; 87779 MW; FEC67F3387228E7F CRC64;

Query Match 92.5%; Score 49; DB 2; Length 805;   
Best Local Similarity 100.0%; Pred. No. 1.9e+02;   
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;   
Qy 2 AAAAAAAAAAAAA 13   
Db 17 AAAAAAAAAAAAA 28

RESULT 26   
ID PYGO DROME STANDARD; PRT; 815 AA.   
AC Q9V9W8;   
DT 19-SEP-2002, integrated into UniProtKB/Swiss-Prot.   
DT 01-MAY-2000, sequence version 1.   
DT 21-FEB-2006, entry version 39.

Qy 1 AAAAAAAAAAAAA 13   
Db 3 AAAAAAAAAAAAA 15

RESULT 25   
ID Q4SPH0\_TETNG PRELIMINARY; PRT; 805 AA.   
AC Q4SPH0;



DE Protein pygopus (Gammy legs protein).  
GN Name=pygo; Synonyms=gam; ORFNames=CG11518;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21952490; PubMed=11953446; DOI=10.1016/S0092-8674(02)00679-7;  
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
RA Murone M., Zuellig S., Basler K.;  
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
RT pygopus to the nuclear beta-catenin-TCF complex.";  
RL Cell 109:47-60(2002).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX TISSUE=Head;  
RA MEDLINE=22010046; PubMed=12015286;  
RA Parker D.S., Jemison J., Cadigan K.M.;  
RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling  
RT in Drosophila.";  
RL Development 129:2565-2576(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX STRAIN=Berkley;  
RA MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe C.R., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Flesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hottin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [4]  
RP GENOME REANNOTATION.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RX STRAIN=Berkley; TISSUE=Embryo;  
RA MEDLINE=22426066; PubMed=12537569;  
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
RA George R.A., Guarin H., Kronmiller B., Pacle B.J.M., Park S., Wan K.H.,  
RA Rubin G.M., Celniker S.E.;  
RT "A Drosophila full-length cDNA resource.";  
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
CC -!- FUNCTION: Involved in signal transduction through the Wnt pathway.  
CC -!- SUBUNIT: Binds to BCL9 via the PHD-type zinc finger motif, and  
CC thereby becomes part of the nuclear ARM/PAN complex.  
CC -!- SUBCELLULAR LOCATION: Nucleus.  
CC -!- TISSUE SPECIFICITY: Ubiquitous throughout embryogenesis and larval  
CC development.  
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
CC throughout development.  
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL: AF457206; AAL91369.1; -; mRNA.  
CC EMBL: AY075095; AAL79357.1; -; mRNA.  
CC EMBL: AY030778; AAF57161.1; -; Genomic DNA.  
CC EMBL: AY058500; AAL13729.1; -; mRNA.  
CC Ensembl: CG11518; Drosophila melanogaster.  
CC FlyBase: FBgn0043900; pygo.  
CC BioCyc: DMEL-XXX-02.DMEL-XXX-02-014325-MONOMER; -.  
CC GO: GO:0005634; C:nucleus; NAS.  
CC GO: GO:0030528; F:transcriptional regulator activity; IPI.  
CC GO: GO:0035214; P:eye-antennal disc development; IMP.  
CC GO: GO:0030177; P:positive regulation of Wnt receptor signaling; IPI.  
CC GO: GO:0007367; P:segment polarity determination; IMP.  
CC GO: GO:0016055; P:Wnt receptor signaling pathway; IMP.  
CC InterPro: IPR001965; Znf\_PHD.  
CC Pfam: PF00628; PHD; 1.  
CC SMART: SM00249; PHD; 1.  
CC PROSITE: PS01359; ZF\_PHD\_1; 1.  
CC PROSITE: PS50016; ZF\_PHD\_2; 1.  
CC Complete proteome; Developmental protein; Metal-binding;  
CC Nuclear protein; Segmentation polarity protein; Wnt signaling pathway;  
CC Zinc; Zinc-finger.  
FT CHAIN 1 815 Protein pygopus.  
FT /FTID=PRO\_0000057124.  
FT ZN FING 747 805 PHD-type.  
FT MOTIF 39 45 Nuclear localization signal (Potential).  
FT COMBIAS 48 65 Ala-rich.  
FT COMBIAS 123 749 Asn/Gly/His/Met/Pro-rich.  
FT CONFLICT 393 393 S -> P (in Ref. 1).  
SQ SEQUENCE 815 AA; 80493 MW; 369FDS45D34BC136 CRC64;  
Query Match 92.5%; Score 49; DB 1; Length 815;  
Best Local Similarity 92.3%; Pred. No. 26+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAAA 13  
DB 53 AAAAAAAAAAAAAA 65  
RESULT 27  
Q3UG42\_MOUSE PRELIMINARY; PRT; 828 AA.  
ID Q3UG42\_MOUSE  
AC Q3UG42;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.



DE B16 F10Y cells cDNA, RIKEN full-length enriched library.  
DE clone:G370031D15 product:SH3 multiple domains 2, full insert sequence.  
GN Name=Sh3md2;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Katayama T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
RA Ambesi-Impombato A., Apweiler R., Aurialy R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Huminicki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Luani S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Motagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petkovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugiuira K., Sultana R., Takenaka Y., Taki K.,  
RA Tannoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Mahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessky C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
[3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RG (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense Transcription in the Mammalian Transcriptome";  
RL Science 309:1564-1566(2005).  
[4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Choithia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmer S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petkovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
[5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bono M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
[6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630(2000).  
[7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J;  
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Yoneda S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Fujisawa S., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771(2000).

(8)
NUCLEOTIDE SEQUENCE.
RP STRAIN=CS7BL/6J;
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Wataniki A.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AK148137; BAE28367.1; -; mRNA.
DR MGI; MGI:1913066; SH3md2.
DR GO; GO:0030027; C:lamellipodium; IDA.
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0046328; P:regulation of JNK cascade; IDA.
DR InterPro; IPR00108; Neu\_cyt\_fact\_2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001841; Znf\_RING.
DR Pfam; PF00018; SH3\_1; 3.
DR Pfam; PF00097; Zf-C3HC4; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 3.
DR SMART; SM00184; RING; 1.
DR SMART; SM00326; SH3; 3.
DR PROSITE; PS00002; SH3; 3.
DR PROSITE; PS00518; ZF\_RING\_1; 1.
DR PROSITE; PS0089; ZF\_RING\_2; 1.
SQ SEQUENCE 828 AA; 86192 MW; 72FE525B907FE0FB CRC64;
Query Match 92.5%; Score 49; DB 2; Length 828;
Best Local Similarity 92.3%; Pred. No. 2e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAAAA 13
DB 419 AAAAAAAAAAAAAA 431
RESULT 28
O70254\_MOUSE PRELIMINARY; PRT; 892 AA.
AC O70254;
DT 01-AUG-1998, integrated into UniProtKB/TrEMBL.
DT 01-AUG-1998, sequence version 1.
DT 07-FEB-2006, entry version 26.
DE Plenty of SH3s.
GN Name=Sh3md2; Synonyms=POSH;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI\_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98151363; PubMed=9482736; DOI=10.1093/emboj/17.5.1395;
RA Tapon N., Nagata K., Lamarche N., Hall A.;
RT "A new rac target POSH is an SH3-containing scaffold protein involved
RT in the JNK and NF-kappaB signalling pathways.";
RL EMBO J. 17:1395-1404 (1998).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AF030131; AAC40070.1; -; mRNA.
DR MGI; MGI:1913066; SH3md2.

DR GO; GO:0030027; C:lamellipodium; IDA.
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0046328; P:regulation of JNK cascade; IDA.
DR InterPro; IPR00108; Neu\_cyt\_fact\_2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001841; Znf\_RING.
DR Pfam; PF00018; SH3\_1; 4.
DR Pfam; PF00097; Zf-C3HC4; 1.
DR PRINTS; PR00499; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 4.
DR SMART; SM00184; RING; 1.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS00002; SH3; 4.
DR PROSITE; PS00518; ZF\_RING\_1; 1.
DR PROSITE; PS0089; ZF\_RING\_2; 1.
SQ SEQUENCE 892 AA; 93435 MW; 5A1030857777B3B0 CRC64;
Query Match 92.5%; Score 49; DB 2; Length 892;
Best Local Similarity 92.3%; Pred. No. 2.1e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAAAAAA 13
DB 419 AAAAAAAAAAAAAA 431
RESULT 28
Q9VE88\_DROME PRELIMINARY; PRT; 897 AA.
AC Q9VE88;
DT 01-MAY-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 2.
DT 21-FEB-2006, entry version 28.
DE CG15803-PA.
GN Name=CG15803; ORFNames=Dmel\_CG15803;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI\_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Ffannkuch C., Baldwin D.,
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Bouch M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22426065; PubMed=12537568;  
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22426070; PubMed=12537573;  
RX Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RN NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RN NUCLEOTIDE SEQUENCE.  
RG Berkeley *Drosophila* Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "*Drosophila melanogaster* release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RN NUCLEOTIDE SEQUENCE.  
RP FlyBase;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC -!- INTERACTION:  
CC QVBP2:CG5053; NbExp=1; IntAct=EBI-172540, EBI-202525;  
CC QVQW7:ed; NbExp=1; IntAct=EBI-172540, EBI-85823;  
CC -!- SIMILARITY: Contains 4 PDZ (DHR) domains.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AE003721; AAF55539.2; -; Genomic\_DNA.  
DR HSSP; O64512; 1021.  
DR IntAct; QVBE88; -.  
DR FlyBase; FBgn0038606; CG15803.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR InterPro; IPR001478; PDZ.  
DR Pfam; PF00595; PDZ; 4.  
DR SMART; SM00228; PDZ; 4.  
DR PROSITE; PS50106; PDZ; 4.

SQ SEQUENCE 897 AA; 94372 MW; 3A74789E08FD1A2F CRC64;  
Query Match 92.5%; Score 49; DB 2; Length 897;  
Best Local Similarity 92.3%; Pred. No. 2.1e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAAA 13  
DB 133 AAAAAAAAAAAAAA 145  
RESULT 30  
O69ZT11\_MOUSE PRELIMINARY; PRT; 914 AA.  
AC O69ZT11;  
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.  
DT 13-SEP-2004, sequence version 1.  
DT 07-FEB-2006, entry version 12.  
DE MKTAA1494 protein (Fragment).  
DE Name=Sh3md2; Synonyms=MKIAA1494;  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Embryonic tail;  
RX PubMed=15368995; DOI=10.1093/dnares/11.3.205;  
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,  
RA Suga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,  
RA Nagase T., Ohara O., Koga H.;  
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries.";  
RL DNA Res. 11:205-218(2004).  
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
CC -!- SIMILARITY: Contains 4 SH3 domains.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AK173185; BAD32463.1; -; mRNA.  
DR MGI; MGI:1913066; Sh3md2.  
DR GO; GO:0030027; C:lamellipodium; IDA.  
DR GO; GO:0005078; F:MAP-kinase scaffold activity; IDA.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR GO; GO:0046328; F:regulation of JNK cascade; IDA.  
DR InterPro; IPR00108; Neu\_cyt\_fact\_2.  
DR InterPro; IPR001452; SH3.  
DR InterPro; IPR001841; Znf\_RING.  
DR Pfam; PF00018; SH3\_1; 4.  
DR Pfam; PF00097; zf-C3HC4; 1.  
DR PRINTS; PR00499; P67PHOX.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 4.  
DR SMART; SM00184; RING; 1.  
DR SMART; SM00326; SH3; 4.  
DR PROSITE; PS50002; SH3; 4.  
DR PROSITE; PS00518; ZF\_RING\_1; 1.  
DR PROSITE; PS0089; ZF\_RING\_2; 1.  
KW Metal-binding; SH3 domain; Zinc; Zinc-finger.  
FT NON\_TER 1  
SQ SEQUENCE 914 AA; 95963 MW; 0D3F7A4B4AD9E067 CRC64;  
Query Match 92.5%; Score 49; DB 2; Length 914;  
Best Local Similarity 92.3%; Pred. No. 2.1e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAAA 13  
DB 441 AAAAAAAAAAAAAA 453

RESULT 31  
 ID Q3UYA4\_MOUSE PRELIMINARY; PRT; 935 AA.  
 AC Q3UYA4;  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 11-OCT-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 5.  
 DE Adult male medulla oblongata cDNA, RIKEN full-length enriched library,  
 DE clone:630596803 product:hypothetical Transforming protein Ski/SAND-  
 DE like/putative DNA binding containing protein, full insert sequence.  
 GN Name=Lbxcor1;  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning."  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,  
 RA Ambesi-Impombato A., Apweiler R., Attalini R.N., Bailey T.L.,  
 RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
 RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
 RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 RA Di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
 RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
 RA Hill D., Hummel S., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
 RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,  
 RA Kitano H., Kollas G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
 RA Kurochkin I.V., Larau L.F., Lazarevic D., Lipovich L., Liu J.,  
 RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
 RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
 RA Mottagui-Tabar S., Mulder N., Nakano N., Nakamura H., Ng P.,  
 RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
 RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
 RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
 RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
 RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
 RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
 RA Sperling K., Stupka E., Suglira K., Sultana R., Takenaka Y., Taki K.,  
 RA Tannoia K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
 RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
 RA Yamashita H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
 RA Grimmond S.M., Teasdale R.D., Liu E.T., Bruscia V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki K., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Komoto H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome."  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group

RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense transcription in the Mammalian Transcriptome."  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikola I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragni T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Iwama M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto P.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes."  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group

RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer.";  
RL Genome Res. 10:1757-1771(2000).  
RN [6]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tegami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AK134840; BAE22308.1; -; mRNA.  
DR MGI; MGI:2443473; Lbxcor1.  
DR GO; GO:0005634; C:nucleus; RCA.  
DR GO; GO:0005634; C:nucleus; RCA.  
DR GO; GO:0005667; C:transcription factor complex; IDA.  
DR GO; GO:0005515; F:protein binding; IPI.  
DR GO; GO:0003714; F:transcription corepressor activity; IDA.  
DR GO; GO:0050875; P:cellular physiological process; RCA.  
DR GO; GO:0016481; P:negative regulation of transcription; IDA.  
DR InterPro: IPR003380; Transform\_Ski.  
DR Pfam: PF02437; Ski\_Sno; 1.  
DR Hypothetical protein.  
KW SEQUENCE 935 AA; 96933 MW; A9C2953F4A77E367 CRC64;  
SQ  
Query Match 92.5%; Score 49; DB 2; Length 935;  
Best Local Similarity 92.3%; Pred. No. 2.2e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 AAAAAAAAAAAAAA 13  
Db 484 AAAAAAAAAAAAAA 496  
RESULT 32  
Q9V193 DROME PRELIMINARY; PRT; 946 AA.  
ID Q9V193 DROME  
AC Q9V193  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 2.  
DT 21-FEB-2006, entry version 31.  
DE CG32466-PA, isoform A.  
GN Name=rn; ORFNames=Dmel\_CG32466;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RN NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Balley R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Krawitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22426065; PubMed=12537568;  
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RN NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RN NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RN NUCLEOTIDE SEQUENCE.  
RP Berkeley Drosophila Genome Project;  
RG Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RN NUCLEOTIDE SEQUENCE.

```

RG FlyBase;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBSJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity)
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE003672; AAF54032.3; -; Genomic_DNA.
DR FlyBase; FBgn0003263; rn.
DR GO; GO:0007456; P:eye development (sensu Endopterygota); IMP.
DR GO; GO:0007480; P:leg morphogenesis (sensu Endopterygota); IMP.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 6.
DR ProDom; PD0000003; Znf_C2H2; 3.
DR SMART; SM00355; Znf_C2H2; 6.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.
SQ SEQUENCE 946 AA; 100948 MW; 05EA01C8F64F6133 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 946;
Best Local Similarity 92.3%; Pred. NO. 2.2e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 275 AAAAAAAAAAAAAA 287

RESULT 33
LXCO MOUSE
ID LXCO_MOUSE STANDARD; PRT; 964 AA.
AC Q8BX46; Q5W812; Q8COT2;
DT 21-JUN-2005, integrated into UniProtKB/Swiss-Prot.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 20.
DE Ladybird homeobox corepressor 1 (Transcriptional corepressor Corl1).
GN Name=Lbxcor1; Synonyms=Corl1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Murioidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 2); FUNCTION, SUBCELLULAR
RP LOCATION, TISSUE SPECIFICITY, AND INTERACTION WITH Lbx1.
RC TISSUE=Embryonic brain;
RX PubMed=1528197; DOI=10.1074/jbc.M411652200;
RA Mizuhara E., Nakatani T., Minaki Y., Sakamoto Y., Ono Y.;
RT "Corl1, a novel neuronal lineage-specific transcriptional corepressor
RL for the homeodomain transcription factor Lbx1.";
RJ J. Biol. Chem. 280:3645-3655(2005).
[2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORMS 1 AND 3).
RP STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama K., Ravasi N., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturalien R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Humnietcki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasaki Y., Kelsø J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,

```

```

RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Motagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sadelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Sempke C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugitara K., Sultana R., Takenaka Y., Taki K.,
RA Tamoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Nimioya N.,
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Wataniki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
CC -!- FUNCTION: Acts as a transcriptional corepressor of Lbx1.
CC -!- SUBUNIT: Interacts with Lbx1.
CC -!- INTERACTION:
CC Q9QVT4;- (xeno); NbExp=1; IntAct=EBI-604451, EBI-604706;
CC Q8712; Ctbp1; NbExp=1; IntAct=EBI-604451, EBI-604547;
CC Q09106; Hdac1; NbExp=1; IntAct=EBI-604451, EBI-301912;
CC P52955; Lbx1; NbExp=1; IntAct=EBI-604451, EBI-604594;
CC Q62440; Tle1; NbExp=1; IntAct=EBI-604451, EBI-604471;
CC -!- SUBCELLULAR LOCATION: Nucleus.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q8BX46-1; Sequence=Displayed;
CC Note=No experimental confirmation available;
CC Name=2;
CC IsoId=Q8BX46-2; Sequence=VSP_014179;
CC Name=3;
CC IsoId=Q8BX46-3; Sequence=VSP_014178, VSP_014180, VSP_014181;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Expressed in brain with higher levels in
CC embryo than adult. Also expressed in adult testis. In embryonic
CC brain, expressed in a subset of postmitotic neurons generated
CC posterior to the midbrain-hindbrain border. In the developing
CC spinal cord, selectively expressed in dorsal horn interneurons.
CC -!- SIMILARITY: Belongs to the SKI family.
-----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AB185113; BAD69568.1; -; mRNA.
DR EMBL; AK029916; BAC26674.1; -; mRNA.
DR EMBL; AK049035; BAC33520.1; -; mRNA.
DR HSP; P12755; 1MR1.
DR IntAct; Q8BX46; -.
DR Ensembl; ENSMUSG00000022245; Mus musculus.
DR MGI; MGI:2443473; Lbxcor1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005667; C:transcription factor complex; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0003714; F:transcription corepressor activity; IDA.
DR GO; GO:0016480; P:negative regulation of transcription; IDA.
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno; 1.
KW Alternative splicing; Coiled coil; Nuclear protein; Repressor;
KW Transcription; Transcription regulation.
CHAIN 1 964 Ladybird homeobox corepressor 1.
/FTId=PRO_0000129391.
COILED 853 921 Potential.
VARSPPLIC 1 143 Missing (in isoform 3).
/FTId=VSP_014178.
VARSPPLIC 9 36 Missing (in isoform 2).

```

```

FT  VARSPLIC 282 290 /FTid=VSP_014179.
FT  LQGGGGGA -> PARCRRRR (in isoform 3).
FT  /FTid=VSP_014180.
FT  VARSPLIC 291 964 Missing (in isoform 3).
FT  /FTid=VSP_014181.
FT  CONFLICT 961 961 Q -> P (in Ref. 1).
FT  SEQUENCE 964 AA; 100276 MW; 100276 MW; BLECDB443D789E2C CRC64;

Query Match 92.5%; Score 49; DB 1; Length 964;
Best Local Similarity 92.3%; Pred. No. 2.2e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
DB 512 AAAAAAAAAAAAAA 524

RESULT 34
LBXCO RAT STANDARD; PRT; 964 AA.
AC P84551;
DT 21-JUN-2005, integrated into UniProtKB/Swiss-Prot.
DT 21-JUN-2005, sequence version 1.
DT 07-MAR-2006, entry version 8.
DE Ladybird homeobox corepressor 1.
GN Name=Lbxcor1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Murioidea; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
NP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Brown Norway;
RX PubMed=15057822; DOI=10.1038/nature02426;
RA Gibbs R.A., Weinstock G.M., Metzger M.L., Muzny D.M., Sodergren E.J.,
RA Scherer S., Scott G., Steffen D., Worley K.C., Burch P.E., Okwuonu G.,
RA Hines S., Lewis L., Dexamio C., Delgado O., Dugan-Rocha S., Miner G.,
RA Morgan M., Hawes A., Gill R., Holt R.A., Adams M.D., Amanatides P.G.,
RA Baden-Tillson H., Barnstead M., Chin S., Evans C.A., Fertiera S.,
RA Foster C., Glodek A., Gu Z., Jennings D., Kraft C.L., Nguyen T.,
RA Pfannkoch C.M., Sitter C., Sutton G.G., Venter J.C., Woodage T.,
RA Smith D., Lee H.M., Gustafson E., Cahill P., Kana A.,
RA Doucette-Stamm L., Weinstock K., Fectel K., De Jong P.J., Osogawa K.,
RA Green E.D., Blakesley R.W., Bouffard G.G., Bosdet I., Fjell C., Jones S.,
RA Zhu B., Marra M., Schein J., Rosdet I., Shatsman S., Geer K., Chen Y.,
RA Krzywinski M., Mathewson S., Siddiqui A., Wye N., McPherson J.,
RA Zhao S., Fraser C.M., Shetty J., Shatsman S., Geer K., Chen Y.,
RA Abramson S., Nierman W.C., Havlak P.H., Chen R., Durbin K.J., Egan A.,
RA Ren Y., Song X.Z., Li B., Liu Y., Qin X., Cawley S., Cooney A.J.,
RA D'Souza L.M., Martin K., Wu J.Q., Gonzalez-Garay M.L., Jackson A.R.,
RA Kalafus K.J., McLeod M.P., Milosavljevic A., Virk D., Volkov A.,
RA Wheeler D.A., Zhang Z., Bailey J.A., Eichler E.E., Tuzun E.,
RA Birney E., Mongin E., Ureta-Vidal A., Woodward C., Zdobnov E.,
RA Bork P., Suyama M., Torrents D., Alexandersson M., Trask B.J.,
RA Young J.M., Huang H., Wang H., Xing H., Daniels S., Gietzen D.,
RA Schmidt J., Stevens K., Vitt U., Wingrove J., Camara F., Mar Alba M.,
RA Abril J.F., Guigo R., Smit A., Dubchak I., Rubin E.M., Couronne O.,
RA Poliakov A., Hubner N., Ganten D., Goesele C., Hummel O., Kreitler T.,
RA Lee Y.A., Monti J., Schulz H., Zimdahl H., Himmelbauer H., Lehrach H.,
RA Jacob H.J., Bromberg S., Gullings-Handley J., Jensen-Seaman M.I.,
RA Kwik A.E., Lazar J., Pasko D., Tonellato P.J., Twigger S.,
RA Ponting C.P., Duarte J.M., Rice S., Goodstadt L., Beato S.A.,
RA Emes R.D., Winter E.E., Webber C., Brandt P., Nyakatura G.,
RA Adatobi M., Chiaromonte F., Elnitski L., Eswara P., Hardison R.C.,
RA Hou M., Kolbe D., Makova K., Miller W., Nekutenko A., Riemer C.,
RA Schwartz S., Taylor J., Yang S., Zhang Y., Lindpaintner K.,
RA Andrews T.D., Caccamo M., Clamp M., Clarke L., Curwen V., Durbin R.,
RA Eyas E., Searle S.M., Cooper G.M., Batzoglou S., Brudno M., Sidow A.,
RA Stone E.A., Payseur B.A., Bourque G., Lopez-Otin C., Puente X.S., V.B.,
RA Chakrabarti K., Chatterji S., Dewey C., Fächter L., Bray N., Yap V.B.,
RA Caspi A., Tesler G., Pevzner P.A., Haussler D., Roskin K.M.,
RA Baertsch R., Clawson H., Furey T.S., Hinrichs A.S., Karolchik D.,

```

```

RA Kent W.J., Rosenbloom K.R., Trumbower H., Weirauch M., Cooper D.N.,
RA Stenson P.D., Ma B., Brent M., Arumugam M., Shteynberg D.,
RA Copley R.R., Taylor M.S., Riethman H., Mudunuri U., Peterson J.,
RA Guyer M., Felsenfeld A., Old S., Mockrin S., Collins F.,
RT "Genome sequence of the Brown Norway rat yields insights into
mammalian evolution."
RL Nature 428:493-521(2004).
CC -!- FUNCTION: Acts as a transcriptional corepressor of LBX1 (By
similarity).
CC -!- SUBUNIT: Interacts with LBX1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nucleus (By similarity).
CC -!- SIMILARITY: Belongs to the SKI family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AABR03062264; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR RGD; 1307687; RGD1307687.
KW Coiled coil; Nuclear protein; Repressor; Transcription;
KW Transcription regulation.
FT CHAIN 1 964 Ladybird homeobox corepressor 1.
FT COILED 857 921 Potential.
FT SEQUENCE 964 AA; 100234 MW; 78D1B922CAB29B7D CRC64;
SQ
Query Match 92.5%; Score 49; DB 1; Length 964;
Best Local Similarity 92.3%; Pred. No. 2.2e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
DB 512 AAAAAAAAAAAAAA 524

RESULT 35
LBXCO HUMAN STANDARD; PRT; 965 AA.
AC P84550;
DT 21-JUN-2005, integrated into UniProtKB/Swiss-Prot.
DT 21-JUN-2005, sequence version 1.
DT 07-MAR-2006, entry version 6.
DE Ladybird homeobox corepressor 1.
GN Name=Lbxcor1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominiidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
NP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP Human chromosome 15 international sequencing consortium;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Acts as a transcriptional corepressor of LBX1 (By
similarity).
CC -!- SUBUNIT: Interacts with LBX1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nucleus (By similarity).
CC -!- SIMILARITY: Belongs to the SKI family.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AC009292; -; NOT_ANNOTATED_CDS; Genomic_DNA.
DR HGNC; HGNC:21326; Lbxcor1.
KW Coiled coil; Nuclear protein; Repressor; Transcription;
KW Transcription regulation.
FT CHAIN 1 965 Ladybird homeobox corepressor 1.
FT COILED 858 922 Potential.
FT SEQUENCE 965 AA; 99831 MW; 313E95B0430A979B CRC64;
SQ
Query Match 92.5%; Score 49; DB 1; Length 965;
Best Local Similarity 92.3%; Pred. No. 2.2e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```



QY 1 AAAAAAAAAAAAA 13  
 DB 515 AAAAAAAAAAAAA 527

RESULT 36  
 Q4SPK5\_TETNG  
 ID Q4SPK5\_TETNG PRELIMINARY; PRT; 1065 AA.  
 AC Q4SPK5;  
 DT 19-JUL-2005, integrated into UniProtKB/TrEMBL.  
 DT 19-JUL-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 3.  
 DE Chromosome 16 SCAP14537, whole genome shotgun sequence.  
 GN ORFNames=GSTENG00014784001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15496914; DOI=10.1038/nature03025;  
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
 RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
 RA Nicaut S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
 RA Biemont C., Skalli Z., Catolico L., Poulain J., De Brardinis V.,  
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,  
 RA Kellis M., Wolff J.-N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.,  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 RT the early vertebrate proto-karyotype.";  
 RL Nature 431:946-957(2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC DR EMBL: CAAB01014537; CAF97427.1; -: Genomic DNA.  
 SQ SEQUENCE 1065 AA; 118021 MW; EDD86A851F655E0F CRC64;  
 Query Match 92.5%; Score 49; DB 2; Length 1065;  
 Best Local Similarity 92.3%; Pred. No. 2.4e+02;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAA 13  
 DB 321 AAAAAAAAAAAAA 333

RESULT 37  
 QBIR69\_DROME  
 ID QBIR69\_DROME PRELIMINARY; PRT; 1181 AA.  
 AC QBIR69;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 19.  
 DE CG1770-PB, isoform B.  
 GN Name=HDAC4; ORFNames=Dmel CG1770;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.C., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacieb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomics perspective.";  
 RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnick S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,



RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RA "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RN Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
[5]  
RN NUCLEOTIDE SEQUENCE.  
RP Berkeley Drosophila Genome Project;  
RG Celnikier S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[6]  
RN NUCLEOTIDE SEQUENCE.  
RP FlyBase;  
RG Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AE003492; AAN093118.1; -; Genomic\_DNA.  
DR FlyBase; FBgn041210; HDAC4.  
DR InterPro; IPR000286; His\_deacetylase.  
DR InterPro; IPR001220; Lectin\_legB.  
DR PANTHER; PTHR10625; His\_deacetylase; 3.  
DR Pfam; PF00850; Hist\_deacetyl1; 1.  
DR PRINTS; PR01270; HDASUPER.  
DR PROSITE; PS00307; LECTIN LEGUME BETA; UNKNOWN 1.  
SQ SEQUENCE 1181 AA; 126044 MW; AD451B406B87I9B1 CRC64;  
  
Query Match 92.5%; Score 49; DB 2; Length 1181;  
Best Local Similarity 92.3%; Pred. No. 2.6e+02;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAAAAAA 13  
Db 423 AAAAAAAAAAAAAA 435  
  
RESULT 38  
Q59E49 DROME PRELIMINARY; PRT; 1199 AA.  
ID Q59E49 DROME  
AC Q59E49;  
DT 26-APR-2005, integrated into UniProtKB/TrEMBL.  
DT 26-APR-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE CGI770-PC, isoform C.  
GN Name=HDAC4; ORFNames=Dmel\_CG1770;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
[12]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22426065; PubMed=12537568;  
RA Celnikier S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
[13]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celnikier S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
[14]  
RN NUCLEOTIDE SEQUENCE.  
RP MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celnikier S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
[15]  
RN NUCLEOTIDE SEQUENCE.  
RP Berkeley Drosophila Genome Project;  
RG Celnikier S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[16]  
RN NUCLEOTIDE SEQUENCE.  
RP FlyBase;  
RG Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AE003492; AAX52490.1; -; Genomic\_DNA.  
DR

```

DR InterPro; IPR000286; His_deacetylase.
DR InterPro; IPR001220; Lectin_legh.
DR Pfam; PTHR10625; His_deacetylase; 3.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
DR PROSITE; PS00307; LECTIN LEGUME BETA; UNKNOWN 1.
SQ SEQUENCE 1199 AA; 128125 MW; 40C97AD33F96169C CRC64;

Query Match 92.5%; Score 49; DB 2; Length 1199;
Best Local Similarity 92.3%; Pred. No. 2.6e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 445 AAAAAAAAAAAAAA 457

RESULT 39
Q819J7 DROME PRELIMINARY; PRT; 1203 AA.
AC Q819J7;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Histone deacetylase dHDAC4 isoform a.
GN Name=HDAC4; Synonyms=HDAC4a; ORFNames=CG1770;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP MEDLINE=22368461; PubMed=12481296; DOI=10.1002/gene.10159;
RA Zeremski M., Stricker J.R., Fischer D., Zusan S.B., Cohen D.;
RT "Histone deacetylase dHDAC4 is involved in segmentation of the
RL Drosophila embryo and is regulated by gap and pair-rule genes.";
RL Genesis 35:31-38(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AF538713; AANI5142.1; -; mRNA.
DR FlyBase; FBgn0041210; HDAC4.
DR InterPro; IPR000286; His_deacetylase.
DR InterPro; IPR001220; Lectin_legh.
DR Pfam; PTHR10625; His_deacetylase; 3.
DR PRINTS; PR01270; HDASUPER.
DR PROSITE; PS00307; LECTIN LEGUME BETA; UNKNOWN 1.
SQ SEQUENCE 1203 AA; 128666 MW; 04B5D0B12088979F CRC64;

Query Match 92.5%; Score 49; DB 2; Length 1203;
Best Local Similarity 92.3%; Pred. No. 2.6e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13
Db 445 AAAAAAAAAAAAAA 457

RESULT 40
Q9VYF3 DROME PRELIMINARY; PRT; 1248 AA.
AC Q9VYF3;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 2.
DT 07-FEB-2006, entry version 26.
DE CG1770-PA, isoform A.
GN Name=HDAC4; ORFNames=Dmel_CG1770;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

```

```

OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
RN [2]
RP MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RL a Genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,

```

```

RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RL systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Ceiniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskaas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC -1- INTERACTION:
CC P29310-2.14-3-3-zeta; NbExp=1; IntAct=EBI-149766, EBI-198120;
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AE003492; AAF48245.2; -; Genomic_DNA.
DR IntAct; Q9VFE3; -
DR FlyBase; FBgn0041210; HDAC4.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR000286; His_deacetylase.
DR InterPro; IPR001220; Lectin_legB.
DR PANTHER; PTHR10625; His_deacetylase; 3.
DR Pfam; PF00850; Hist_deacetyl; 1.
DR PRINTS; PR01270; HDASUPER.
DR PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
DR SEQUENCE 1248 AA; 133650 MW; D300C8523CA2C311 CRC64;

Query Match 92.5%; Score 49; DB 2; Length 1248;
Best Local Similarity 92.3%; Pred. No. 2.7e+02;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
Db 494 AAAAAAAAAAAAAA 506

```

Search completed: September 9, 2006, 23:00:20  
Job time : 125.139 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:37:17 ; Search time 97.4177 Seconds  
(without alignments)  
61.014 Million cell updates/sec

Title: US-10-617-568-4

Perfect score: 54

Sequence: 1 AAFAAAAAAAAA 13

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A\_Geneseq\_8.\*

1: geneseqp1980s.\*

2: geneseqp1990s.\*

3: geneseqp2000s.\*

4: geneseqp2001s.\*

5: geneseqp2002s.\*

6: geneseqp2003as.\*

7: geneseqp2003bs.\*

8: geneseqp2004s.\*

9: geneseqp2005s.\*

10: geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	13	ADI29008	ADI29008 Human MHC
2	51	94.4	15	ABP52215	ABP52215 HLA-DR2 m
3	49	90.7	13	ADW35574	ADW35574 HLA bindi
4	49	90.7	13	ADW34341	ADW34341 HLA bindi
5	49	90.7	14	ADW35752	ADW35752 HLA bindi
6	49	90.7	14	ADW34519	ADW34519 HLA bindi
7	49	90.7	422	ABP70644	ABP70644 Drosophil
8	49	90.7	465	ADW75312	ADW75312 Prostate
9	48	88.9	13	ADI29007	ADI29007 Human MHC
10	48	88.9	278	ADR98917	ADR98917 Lung spec
11	48	88.9	285	ABR65204	ABR65204 Drosophil
12	48	88.9	646	ADY09213	ADY09213 Plant ful
13	48	88.9	1444	ABG15667	ABG15667 Novel hum
14	47	87.0	15	3 AAY58983	AAY58983 Copeptide
15	47	87.0	15	3 AAY58972	AAY58972 Copeptide
16	47	87.0	15	3 AAY58985	AAY58985 Copeptide
17	47	87.0	15	3 AAY82075	AAY82075 MHC class
18	47	87.0	15	3 AAY82061	AAY82061 MHC class
19	47	87.0	15	3 AAY82077	AAY82077 MHC class
20	47	87.0	15	4 AAG63197	AAG63197 Peptide w
21	47	87.0	15	4 AAG63203	AAG63203 Peptide w
22	47	87.0	15	4 AAG63194	AAG63194 Peptide w
23	47	87.0	15	5 ABP52216	ABP52216 HLA-DR2 m

24	47	87.0	15	5 ABP52286	ABP52286 HLA-DR2 m
25	47	87.0	15	5 ABP52212	ABP52212 HLA-DR2 m
26	47	87.0	15	5 ABP52214	ABP52214 HLA-DR2 m
27	47	87.0	15	5 ABP52210	ABP52210 HLA-DR2 m
28	47	87.0	15	7 ADJ57538	ADJ57538 Cop-1 rel
29	47	87.0	15	7 ADJ57547	ADJ57547 Cop-1 rel
30	47	87.0	15	7 ADJ57541	ADJ57541 Cop-1 rel
31	47	87.0	15	8 ADQ59600	ADQ59600 Copolymer
32	47	87.0	15	8 ADQ59606	ADQ59606 Copolymer
33	47	87.0	15	8 ADQ59597	ADQ59597 Copolymer
34	47	87.0	15	8 ADQ81224	ADQ81224 Copolymer
35	47	87.0	15	8 ADQ81233	ADQ81233 Copolymer
36	47	87.0	15	8 ADQ81227	ADQ81227 Copolymer
37	47	87.0	15	9 ADW86548	ADW86548 Ordered c
38	47	87.0	15	9 ADW86539	ADW86539 Ordered c
39	47	87.0	15	9 ADW86542	ADW86542 Ordered c
40	47	87.0	15	9 AEA35261	AEA35261 Cop 1 bin
41	47	87.0	15	9 AEA35270	AEA35270 Cop 1 bin
42	47	87.0	15	9 AEA35264	AEA35264 Cop 1 bin
43	47	87.0	15	9 AEA53005	AEA53005 COP-1 pep
44	47	87.0	15	9 AEA53002	AEA53002 COP-1 pep
45	47	87.0	15	9 AEA53011	AEA53011 COP-1 pep
46	47	87.0	15	9 AEE20290	AEE20290 Random co
47	47	87.0	15	9 AEE20293	AEE20293 Random co
48	47	87.0	15	9 AEE20299	AEE20299 Random co
49	47	87.0	15	10 AEF02691	AEF02691 Random co
50	47	87.0	15	10 AEF02689	AEF02689 Random co
51	47	87.0	15	10 AEF02688	AEF02688 Random co
52	47	87.0	15	10 AEF87650	AEF87650 Autoantig
53	47	87.0	15	10 AEF87641	AEF87641 Autoantig
54	47	87.0	15	10 AEF87644	AEF87644 Autoantig
55	47	87.0	109	2 AAR84086	AAR84086 T-lymphoc
56	47	87.0	373	4 ABO58367	ABO58367 Human gen
57	47	87.0	112	4 ABO15501	ABO15501 Novel hum
58	47	87.0	252	6 AAO16493	AAO16493 Zorocrate
59	47	87.0	252	6 AAE36894	AAE36894 Zorocrate
60	47	87.0	316	4 ABO18917	ABO18917 Novel hum
61	47	87.0	373	6 AAB69577	AAB69577 Drosophil
62	47	87.0	373	6 AAE35778	AAE35778 Human SEC
63	47	87.0	378	5 ABP65077	ABP65077 Hypoxia-i
64	47	87.0	393	4 AAM25623	AAM25623 Human pro
65	47	87.0	394	6 AAE35774	AAE35774 Human SEC
66	47	87.0	512	7 ADB64824	ADB64824 Human pro
67	47	87.0	578	4 ABB59750	ABB59750 Drosophil
68	47	87.0	708	4 AAB92956	AAB92956 Human pro
69	47	87.0	815	4 ABB70439	ABB70439 Drosophil
70	47	87.0	815	6 ABP54340	ABP54340 Drosophil
71	47	87.0	855	4 ABB71577	ABB71577 Drosophil
72	47	87.0	892	7 ABU64278	ABU64278 Murine ub
73	47	87.0	892	7 ADH73564	ADH73564 Mouse POS
74	47	87.0	892	8 ADR89697	ADR89697 Murine PO
75	47	87.0	892	8 ADS34245	ADS34245 POSH prot
76	47	87.0	892	8 ADU77904	ADU77904 Mouse POS
77	47	87.0	892	8 ADU73761	ADU73761 Mouse POS
78	47	87.0	892	8 ADU68997	ADU68997 House mou
79	47	87.0	892	9 ADW87412	ADW87412 Mouse POS
80	47	87.0	892	9 ADZ66429	ADZ66429 Mouse POS
81	47	87.0	892	9 AEA10631	AEA10631 Murine PO
82	47	87.0	892	9 AEC94736	AEC94736 Mouse POS
83	47	87.0	892	4 ABB66003	ABB66003 Drosophil
84	47	87.0	1184	4 ABB58493	ABB58493 Drosophil
85	47	87.0	1235	4 ABB58985	ABB58985 Drosophil
86	47	87.0	1337	4 ABB58985	ABB58985 Drosophil
87	47	87.0	1344	6 AAE36107	AAE36107 Human chr
88	47	87.0	1944	7 ADI26773	ADI26773 Human CHD
89	47	87.0	1944	7 ADJ68739	ADJ68739 Human hea
90	47	87.0	1944	7 ADJ68739	ADJ68739 Human hea
91	47	87.0	2000	7 ADI26747	ADI26747 Human GS-
92	46	85.2	13	2 AAR74159	AAR74159 Antigenic
93	46	85.2	13	2 AAR87437	AAR87437 Human MHC
94	46	85.2	13	3 AAY99164	AAY99164 HLA class
95	46	85.2	13	3 AAY99163	AAY99163 HLA class
96	46	85.2	13	3 AAY99245	AAY99245 HLA class

97 46 85.2 13 7 ADM35620 HLA bindi  
 98 46 85.2 13 7 ADM34387 HLA bindi  
 99 46 85.2 13 8 ADM64988 HLA bindi  
 100 46 85.2 13 9 ADV78640 Cell atta

## ALIGNMENTS

RESULT 1  
 ADI29008  
 ID ADI29008 standard; peptide; 13 AA.  
 XX  
 AC ADI29008;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Human MHC class II compound spaceholder molecule SEQ ID NO:4.  
 XX  
 KW MHC class II compound; MHC class II component; MHC class II alpha chain;  
 KW MHC class II beta chain; peptide binding groove; spaceholder molecule;  
 KW effector component; immune response; immune disorder; virucide;  
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
 KW gene therapy; viral infections; bacterial infection; parasitic infection;  
 KW neoplastic disease; autoimmunity; toxicity; human.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO2004007528-A2.  
 XX  
 PD 22-JAN-2004.  
 XX  
 XX 11-JUL-2003; 2003WO-US021767.  
 XX  
 PF 12-JUL-2002; 2002US-0395494P.  
 PR 22-JUL-2002; 2002US-0397893P.  
 XX  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 XX  
 XX Wucherpfenning KW, Seth N;  
 PI WPI; 2004-122876/12.  
 DR  
 XX  
 XX New MHC class II compound, useful for preparing a composition for  
 PT treating immune disorders e.g. viral infections, bacterial infections,  
 PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
 XX  
 XX Claim 11; SEQ ID NO 4; 92pp; English.  
 PS  
 XX The present invention describes an isolated MHC class II compound (1)  
 CC comprising: (a) an MHC class II component comprising at least a portion  
 CC of an MHC class II alpha chain and at least a portion of an MHC class II  
 CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
 CC chain form a peptide binding groove; (b) a spaceholder molecule; and (c)  
 CC an effector component, where the effector component is linked to the MHC  
 CC class II component. Also described: (1) a pharmaceutical composition  
 CC comprising the MHC class II molecule and a carrier; (2) a method of  
 CC producing an MHC class II compound; (3) a method of directly identifying  
 CC an antigen-specific T cell; (4) a method of regulating an immune response  
 CC in a subject; (5) a method of treating an immune disorder in a subject;  
 CC (6) a method of regulating an immune response ex vivo in a subject; and  
 CC (7) a method of treating an immune disorder ex vivo in a subject. (1) has  
 CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive  
 CC activities, and can be used in gene therapy. The MHC class II compound  
 CC (1) can be used for preparing a composition for treating immune  
 CC disorders, e.g., viral infections, bacterial infections, parasitic  
 CC infections, neoplastic disease, autoimmunity or toxicity. The present  
 CC sequence represents a spaceholder molecule peptide, which can be used in  
 CC an MHC class II compound from the present invention.  
 XX  
 XX Sequence 13 AA;

Query Match 100.0%; Score 54; DB 8; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAFAAAAAAA 13  
 |||||  
 DB 1 AAFAAAAAAA 13  
 |||||  
 RESULT 2  
 ABP52215  
 ID ABP52215 standard; peptide; 15 AA.  
 XX  
 AC ABP52215;  
 XX  
 DT 16-OCT-2002 (first entry)  
 XX  
 DE HLA-DR2 molecule binding peptide SEQ ID NO:9.  
 XX  
 KW Human leukocyte antigen; HLA; major histocompatibility complex; MHC;  
 KW HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;  
 KW immune response; antiinflammatory; neuroprotective; proliferation;  
 KW MHC class II protein inhibitor; demyelinating disease; inhibition;  
 KW post-viral encephalomyelitis; post-vaccine demyelinating condition;  
 KW anti-tumour necrosis factor agent.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200259143-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 XX 24-JAN-2002; 2002WO-US002071.  
 XX  
 PR 24-JAN-2001; 2001US-0263569P.  
 XX  
 XX (HARD ) HARVARD COLLEGE.  
 PA  
 XX Strominger JL, Fridkis-Hareli M;  
 PI WPI; 2002-608439/65.  
 DR  
 XX New compositions comprising synthetic peptides in complex with a major  
 PT histocompatibility complex class II HLA-DR2 protein, useful for treating a  
 PT demyelinating disease, e.g. multiple sclerosis, or post-viral  
 PT encephalomyelitis.  
 XX  
 XX Example 1; Page 30; 54pp; English.  
 PS  
 XX The present invention describes compositions (I) comprising a peptide  
 CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine  
 CC residues. The complex of the peptide with a major histocompatibility  
 CC complex (MHC) class II HLA-DR2 protein is involved in modulating an  
 CC immune response. (I) has antiinflammatory and neuroprotective activities,  
 CC and can be used as a MHC class II protein inhibitor. The compositions  
 CC comprising the peptides are useful for treating demyelinating diseases  
 CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine  
 CC demyelinating condition, and a side effect of administering an anti-  
 CC tumour necrosis factor agents. The peptide further inhibits proliferation  
 CC of autoantigen-specific HLA-DR2-restricted T cell clones. ABP52207 to  
 CC ABP52305 represent peptides used in the exemplification of the present  
 CC invention  
 XX  
 XX Sequence 15 AA;  
 SQ  
 Query Match 94.4%; Score 51; DB 5; Length 15;  
 Best Local Similarity 92.3%; Pred. No. 0.42;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAFAAAAAAA 13  
 |||||  
 DB 3 AAYAAAAAAA 15  
 |||||

```

RESULT 3
ADW35574
ID ADW35574 standard; peptide; 13 AA.
XX AC ADW35574;
XX
DT 10-MAR-2005 (first entry)
XX
DE HLA binding epitope #6324.
XX
KW Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;
KW viral disease; cancer.
XX
OS Unidentified.
XX
PN WO2003040165-A2.
XX
PD 15-MAY-2003.
XX
PF 18-OCT-2001; 2001WO-US051650.
XX
PR 19-OCT-2000; 2000US-0242350P.
XX
PR 20-APR-2001; 2001US-0285624P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S;
XX
DR WPI; 2003-441519/41.
XX
PT New composition comprising at least one peptide having allele-specific
PT binding motifs for HLA, useful for preventing, treating or diagnosing
PT viral diseases and cancer.
XX
PS Claim 1; Page 52-379; 382pp; English.
XX
CC The invention relates to a composition comprising at least one peptide
CC having an isolated, prepared epitope selected from any of the sequences
CC from 30 lists given in the specification. Also disclosed is a method for
CC inducing a cytotoxic T cell response against a pre-selected antigen in a
CC patient expressing a specific MHC class I allele by contacting cytotoxic
CC T cells from the patient with the composition cited above. The
CC composition comprises an epitope that is joined by an amino acid linker.
CC The epitope is admixed or joined to a CTL or HTL epitope. The epitope is
CC bound to an HLA molecule on the antigen-presenting cell, where when an A2
CC -restricted cytotoxic lymphocyte (CTL) is present, a receptor of the CTL
CC binds to a complex of the HLA molecule and the epitope. Specifically
CC claimed are peptides having allele-specific binding motifs for HLA. The
CC compositions and methods are useful for preventing, treating or
CC diagnosing viral diseases and cancer. The peptide epitopes are useful as
CC diagnostic agents for evaluating immune responses, for making antibodies
CC and for evaluating efficacy of a vaccine. Sequences given in ADW29251-
CC ADW37745 represent epitopes of the invention as given in Tables 2-31.
XX
SQ Sequence 13 AA;
XX
Query Match 90.7%; Score 49; DB 7; Length 13;
Best Local Similarity 92.3%; Pred. No. 0.71;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
DQ 1 AAFAAAAA 13
XX
RESULT 4
ADW34341
ID ADW34341 standard; peptide; 13 AA.
XX AC ADW34341;
XX
DT 10-MAR-2005 (first entry)
XX
DE HLA binding epitope #6324.
XX
KW Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;
KW viral disease; cancer.
XX
OS Unidentified.
XX
PN WO2003040165-A2.
XX
PD 15-MAY-2003.
XX
PF 18-OCT-2001; 2001WO-US051650.
XX
PR 19-OCT-2000; 2000US-0242350P.
XX
PR 20-APR-2001; 2001US-0285624P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S;
XX
DR WPI; 2003-441519/41.
XX
PT New composition comprising at least one peptide having allele-specific
PT binding motifs for HLA, useful for preventing, treating or diagnosing
PT viral diseases and cancer.
XX
PS Claim 1; Page 52-379; 382pp; English.
XX
CC The invention relates to a composition comprising at least one peptide
CC having an isolated, prepared epitope selected from any of the sequences
CC from 30 lists given in the specification. Also disclosed is a method for
CC inducing a cytotoxic T cell response against a pre-selected antigen in a
CC patient expressing a specific MHC class I allele by contacting cytotoxic
CC T cells from the patient with the composition cited above. The
CC composition comprises an epitope that is joined by an amino acid linker.
CC The epitope is admixed or joined to a CTL or HTL epitope. The epitope is
CC bound to an HLA molecule on the antigen-presenting cell, where when an A2
CC -restricted cytotoxic lymphocyte (CTL) is present, a receptor of the CTL
CC binds to a complex of the HLA molecule and the epitope. Specifically
CC claimed are peptides having allele-specific binding motifs for HLA. The
CC compositions and methods are useful for preventing, treating or
CC diagnosing viral diseases and cancer. The peptide epitopes are useful as
CC diagnostic agents for evaluating immune responses, for making antibodies
CC and for evaluating efficacy of a vaccine. Sequences given in ADW29251-
CC ADW37745 represent epitopes of the invention as given in Tables 2-31.
XX
SQ Sequence 13 AA;
XX
Query Match 90.7%; Score 49; DB 7; Length 13;
Best Local Similarity 92.3%; Pred. No. 0.71;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
DQ 1 AAFAAAAA 13
XX
RESULT 5
ADW35752
ID ADW35752 standard; peptide; 14 AA.
XX AC ADW35752;
XX
DT 10-MAR-2005 (first entry)
XX
DE HLA binding epitope #6502.
XX
KW Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;

```

```

XX
DT 10-MAR-2005 (first entry)
XX
DE HLA binding epitope #5091.
XX
KW Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;
KW viral disease; cancer.
XX
OS Unidentified.
XX
PN WO2003040165-A2.
XX
PD 15-MAY-2003.
XX
PF 18-OCT-2001; 2001WO-US051650.
XX
PR 19-OCT-2000; 2000US-0242350P.
XX
PR 20-APR-2001; 2001US-0285624P.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Sette A, Sidney J, Southwood S;
XX
DR WPI; 2003-441519/41.
XX
PT New composition comprising at least one peptide having allele-specific
PT binding motifs for HLA, useful for preventing, treating or diagnosing
PT viral diseases and cancer.
XX
PS Claim 1; Page 52-379; 382pp; English.
XX
CC The invention relates to a composition comprising at least one peptide
CC having an isolated, prepared epitope selected from any of the sequences
CC from 30 lists given in the specification. Also disclosed is a method for
CC inducing a cytotoxic T cell response against a pre-selected antigen in a
CC patient expressing a specific MHC class I allele by contacting cytotoxic
CC T cells from the patient with the composition cited above. The
CC composition comprises an epitope that is joined by an amino acid linker.
CC The epitope is admixed or joined to a CTL or HTL epitope. The epitope is
CC bound to an HLA molecule on the antigen-presenting cell, where when an A2
CC -restricted cytotoxic lymphocyte (CTL) is present, a receptor of the CTL
CC binds to a complex of the HLA molecule and the epitope. Specifically
CC claimed are peptides having allele-specific binding motifs for HLA. The
CC compositions and methods are useful for preventing, treating or
CC diagnosing viral diseases and cancer. The peptide epitopes are useful as
CC diagnostic agents for evaluating immune responses, for making antibodies
CC and for evaluating efficacy of a vaccine. Sequences given in ADW29251-
CC ADW37745 represent epitopes of the invention as given in Tables 2-31.
XX
SQ Sequence 13 AA;
XX
Query Match 90.7%; Score 49; DB 7; Length 13;
Best Local Similarity 92.3%; Pred. No. 0.71;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAFAAAAA 13
DQ 1 AAFAAAAA 13
XX
RESULT 5
ADW35752
ID ADW35752 standard; peptide; 14 AA.
XX AC ADW35752;
XX
DT 10-MAR-2005 (first entry)
XX
DE HLA binding epitope #6502.
XX
KW Virucide; cytostatic; gene therapy; vaccine; epitope; cytotoxic T cell;
KW MHC class I; CTL; HTL; A2-restricted cytotoxic lymphocyte; HLA;

```



PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL114747.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 38724; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 422 AA;  
 SQ

Query Match 90.7%; Score 49; DB 4; Length 422;  
 Best Local Similarity 92.3%; Pred. No. 17;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
 | | | | | | | | | |  
 Db 65 AAAAAAAAAAAAAA 77

RESULT 8  
 ADB75312  
 ID ADB75312 standard; protein; 465 AA.  
 XX  
 AC ADB75312;  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Prostate cancer marker protein.  
 KW Prostate; cancer; cytostatic; gene therapy; marker.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO2003009814-A2.  
 PN  
 XX 06-FEB-2003.  
 PD  
 XX 25-JUL-2002; 2002WO-US023913.  
 PF  
 XX 25-JUL-2001; 2001US-0307982P.  
 PR 22-AUG-2001; 2001US-0314356P.  
 PR 25-SEP-2001; 2001US-0325020P.  
 PR 12-DEC-2001; 2001US-0341746P.  
 PR 05-MAR-2002; 2002US-0362158P.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 XX Schlegel R, Monahan JE, Endege WO, Gannavarapu M, Gorbacheva B;  
 PI Hoersht S, Kamatkar S, Wonsey AM, Glatt K, Zhao X, Anderson D;  
 XX  
 XX WPI; 2003-248033/24.  
 DR  
 XX New nucleic acid molecule, useful for diagnosing or treating prostate  
 PT cancer.  
 PT  
 XX Disclosure; SEQ ID NO 136; 99pp; English.  
 PS  
 XX The invention relates to newly discovered cancer markers associated with  
 CC the cancerous state of prostate cells. Also disclosed is a method of

CC assessing whether a patient is afflicted with prostate cancer. The method  
 CC of the invention involves assessing whether a patient is afflicted with  
 CC prostate cancer by comparing the level of expression of a marker in a  
 CC patient sample and the normal level of expression of the marker in a  
 CC control non-prostate cancer sample, where a significant increase in the  
 CC level of expression of the marker in the patient sample and the normal  
 CC level indicates that the patient is afflicted with prostate cancer.  
 CC Nucleic acids of the invention are useful for diagnosing or treating  
 CC prostate cancer, and may be useful in gene therapy. Sequences given in  
 CC ADB75177-ADB75631 represent marker cDNA and proteins. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 465 AA;  
 SQ

Query Match 90.7%; Score 49; DB 7; Length 465;  
 Best Local Similarity 92.3%; Pred. No. 19;  
 Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAAA 13  
 | | | | | | | | | |  
 Db 290 ALFAAAAAAAAAAAAA 302

RESULT 9  
 ADI29007  
 ID ADI29007 standard; peptide; 13 AA.  
 XX  
 AC ADI29007;  
 XX  
 DT 15-APR-2004 (first entry)  
 XX  
 DE Human MHC class II compound spaceholder molecule SEQ ID NO:3.  
 XX  
 KW MHC class II compound; MHC class II component; MHC class II alpha chain;  
 KW MHC class II beta chain; peptide binding groove; spaceholder molecule;  
 KW effector component; immune response; immune disorder; virucide;  
 KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
 KW gene therapy; viral infections; bacterial infection; parasitic infection;  
 KW neoplastic disease; autoimmunity; toxicity; human.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX WO2004007528-A2.  
 PN  
 XX 22-JAN-2004.  
 PD  
 XX 11-JUL-2003; 2003WO-US021767.  
 PF  
 XX 12-JUL-2002; 2002US-0395494P.  
 PR 22-JUL-2002; 2002US-0397893P.  
 PR  
 XX (DAND ) DANA FARBER CANCER INST INC.  
 XX  
 XX Wucherpfenning KW, Seth N;  
 PI  
 XX WPI; 2004-122876/12.  
 DR  
 XX New MHC class II compound, useful for preparing a composition for  
 PT treating immune disorders e.g. viral infections, bacterial infections,  
 PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
 XX  
 XX Claim 11; SEQ ID NO 3; 92pp; English.  
 PS  
 XX The present invention describes an isolated MHC class II compound (I)  
 CC comprising: (a) an MHC class II component comprising at least a portion  
 CC of an MHC class II alpha chain and at least a portion of an MHC class II  
 CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
 CC chain form a peptide binding groove; (b) a spaceholder molecule; and (c)  
 CC an effector component, where the effector component is linked to the MHC  
 CC class II component. Also described: (1) a pharmaceutical composition





SQ Sequence 285 AA;  
Query Match 88.9%; Score 48; DB 4; Length 285;  
Best Local Similarity 92.3%; Pred. No. 17;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAFMAAAAAAAAAA 13  
Db 189 AAAAAAAAAAAAAA 201  
RESULT 12  
ADY09213  
ID ADY09213 standard; protein; 646 AA.  
XX AC ADY09213;  
XX DT 21-APR-2005 (first entry)  
XX DE Plant full length insert polypeptide seqid 65028.  
XX KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content.  
XX OS Unidentified.  
XX US2004034888-A1.  
XX PN 19-FEB-2004.  
XX PD 28-APR-2003; 2003US-00425114.  
XX PF 06-MAY-1999; 99US-00304517.  
XX PR 05-NOV-2001; 2001US-00985678.  
XX PA (LIUJ/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TABA/) TABASKA J E.  
PA (CAOY/) CAO Y.  
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX Claim 1; SEQ ID NO 65028; 15pp; English.  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.seqdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one

CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.  
XX SQ Sequence 646 AA;  
Query Match 88.9%; Score 48; DB 8; Length 646;  
Best Local Similarity 92.3%; Pred. No. 36;  
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAFMAAAAAAAAAA 13  
Db 4 AAAAAAAAAAAAAA 16  
RESULT 13  
ABG15667  
ID ABG15667 standard; protein; 1444 AA.  
XX AC ABG15667;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #15658.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX WO200175067-A2.  
XX PN 11-OCT-2001.  
XX PD 30-MAR-2001; 2001WO-US008631.  
XX PF 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX DR N-PSDB; AAS79854.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX Claim 20; SEQ ID NO 46026; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in



PT aliphatic and the other is charged, used to treat autoimmune diseases.  
 XX  
 PS Claim 82; Page 88; 147pp; English.  
 XX  
 CC The present sequence represents a copeptide that can be used to treat  
 CC autoimmune conditions. It is an example of copeptides of the invention  
 CC that were designed to bind to the groove of human leukocyte antigen HLA-  
 CC DR1 and HLA-DR4 molecules. The copeptides show a high affinity for major  
 CC histocompatibility complex (MHC) class II proteins associated with an  
 CC autoimmune disease, especially HLA-DR1, HLA-DR2 or HLA-DR4, bind to  
 CC antigen presenting cells, and inhibit T cell responses. They are  
 CC especially used to treat multiple sclerosis, autoimmune haemolytic  
 CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, chronic immune thrombocytopenic purpura, colitis, contact  
 CC sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's  
 CC syndrome, Hashimoto's disease, idiopathic myxedema, myasthenia gravis,  
 CC psoriasis, pemphigus vulgaris, rheumatoid arthritis and systemic lupus  
 CC erythematosus (all claimed). The present peptide is a high level  
 CC inhibitor of both HLA-DR-1 and -DR-4-restricted type II collagen-specific  
 CC T cells  
 XX  
 SQ Sequence 15 AA;  
 Query Match 87.0%; Score 47; DB 3; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 AFASAAAAAAA 13  
 |:|||||  
 DB 4 AYASAAAAAAA 15  
 |:|||||  
 RESULT 16  
 AAY58985  
 ID AAY58985 standard; peptide; 15 AA.  
 XX  
 AC AAY58985;  
 XX  
 XX 23-MAY-2000 (first entry)  
 DT  
 XX  
 DE Copeptide useful for treatment of autoimmune disease.  
 XX  
 KW Copeptide; copolymer 1; autoimmune disease; multiple sclerosis;  
 KW autoimmune haemolytic anaemia; autoimmune oophoritis;  
 KW autoimmune thyroiditis; autoimmune uveoretinitis;  
 KW chronic immune thrombocytopenic purpura; colitis;  
 KW contact sensitivity disease; diabetes mellitus; Graves disease;  
 KW Guillain-Barre's syndrome; Hashimoto's disease; idiopathic myxedema;  
 KW myasthenia gravis; psoriasis; pemphigus vulgaris; rheumatoid arthritis;  
 KW systemic lupus erythematosus; immunosuppressant; neuroprotective;  
 KW antianaemic; antithyroid; antidiabetic; thyromimetic; antipsoriatic;  
 KW antirheumatic; antiarthritic; dermatological; antiinflammatory; therapy;  
 KW major histocompatibility complex; MHC class II; human lymphocyte antigen;  
 KW HLA-DR.  
 XX  
 OS Synthetic.  
 XX  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 4 /note="may be substituted by Val or Phe"  
 FT  
 XX  
 XX WO200005250-A1.  
 XX  
 PD 03-FEB-2000.  
 XX  
 XX 23-JUL-1999; 99WO-US016747.  
 XX  
 XX 23-JUL-1998; 98US-0093859P.  
 PR 25-SEP-1998; 98US-0101825P.  
 PR 02-OCT-1998; 98US-0102960P.  
 PR 12-NOV-1998; 98US-0108184P.  
 PR 09-MAR-1999; 99US-0123675P.  
 XX

PA (YEDA ) YEDA RES & DEV CO LTD.  
 PA (HARD ) HARVARD COLLEGE.  
 XX  
 PI Aharoni R, Teitelbaum D, Arnon R, Sela M, Fridkis-Hareli M;  
 XX Strominger JL;  
 DR WPI; 2000-182641/16.  
 XX  
 PT New terpolymers, copeptides and copolymer 1 which contain three amino  
 PT acids randomly joined in a linear array where one is aromatic, one is  
 PT aliphatic and the other is charged, used to treat autoimmune diseases.  
 XX  
 PS Claim 137; Page 69; 147pp; English.  
 XX  
 CC The present sequence represents a copeptide that can be used to treat  
 CC autoimmune conditions. It is an example of copeptides of the invention  
 CC that were designed to bind to the groove of human leukocyte antigen HLA-  
 CC DR1 and HLA-DR4 molecules. The copeptides show a high affinity for major  
 CC histocompatibility complex (MHC) class II proteins associated with an  
 CC autoimmune disease, especially HLA-DR1, HLA-DR2 or HLA-DR4, bind to  
 CC antigen presenting cells, and inhibit T cell responses. They are  
 CC especially used to treat multiple sclerosis, autoimmune haemolytic  
 CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune  
 CC uveoretinitis, chronic immune thrombocytopenic purpura, colitis, contact  
 CC sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's  
 CC syndrome, Hashimoto's disease, idiopathic myxedema, myasthenia gravis,  
 CC psoriasis, pemphigus vulgaris, rheumatoid arthritis and systemic lupus  
 CC erythematosus (all claimed)  
 XX  
 SQ Sequence 15 AA;  
 Query Match 87.0%; Score 47; DB 3; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 AFASAAAAAAA 13  
 |:|||||  
 DB 3 AYASAAAAAAA 14  
 |:|||||  
 RESULT 17  
 AAY82075  
 ID AAY82075 standard; peptide; 15 AA.  
 XX  
 AC AAY82075;  
 XX  
 DT 01-JUN-2000 (first entry)  
 XX  
 DE MHC class II protein binding peptide SEQ ID NO:24.  
 XX  
 KW MHC class II; major histocompatibility complex; autoimmune disease;  
 KW inflammatory disease; binding; rheumatoid arthritis; antiinflammatory;  
 KW antiarthritic; multiple sclerosis.  
 XX  
 OS Synthetic.  
 XX  
 XX WO200005249-A2.  
 XX  
 PD 03-FEB-2000.  
 XX  
 XX 22-JUL-1999; 99WO-US016617.  
 XX  
 PR 23-JUL-1998; 98US-0093859P.  
 PR 09-MAR-1999; 99US-0123675P.  
 XX  
 XX (HARD ) HARVARD COLLEGE.  
 XX  
 XX Strominger JL, Fridkis-Hareli M;  
 XX WPI; 2000-205374/18.  
 XX  
 PT New synthetic peptide, useful for treating autoimmune disease, e.g.  
 PT rheumatoid arthritis.

XX Example 8; Page 32; 57pp; English.

XX The present invention describes synthetic peptides having an amino acid

CC sequence comprising at least 3 residues selected from the group of amino

CC acids consisting of aromatic acids, negatively charged amino acids,

CC positively charged amino acids, and aliphatic amino acids, the synthetic

CC peptides being at least 7 amino acid residues in length and capable of

CC binding to a major histocompatibility complex (MHC) class II protein

CC associated with an autoimmune disease. The synthetic peptides have anti-

CC inflammatory and anti-arthritis activities. They are used to treat

CC inflammatory and demyelinating autoimmune diseases, especially rheumatoid

CC arthritis and multiple sclerosis. The peptides are specific for

CC particular MHC Class II alleles. Purified, short and synthetic peptides

CC should have fewer side effects than mixtures of random peptides; may

CC include many repeats of the active sequence and/or contain amino acid

CC analogues that improve stability (or other desired features). AAY82021 to

CC AAY82044 represent specifically claimed peptide sequences which can be

CC used as part of the synthetic peptides of the present invention; AAY82045

CC to AAY82063 represent specifically claimed examples of the synthetic

CC peptides from the present invention; and AAY82064 to AAY82080 represent

CC other peptides used in the exemplification of the present invention

XX

SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 3; Length 15;

Best Local Similarity 91.7%; Pred. No. 1.6;

Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13

Db 3 AYASAAAAAAA 14

## RESULT 18

AAY82061

ID AAY82061 standard; peptide; 15 AA.

XX

AC AAY82061;

XX

DT 01-JUN-2000 (first entry)

XX

DE MHC class II protein binding peptide SEQ ID NO:33.

XX

KW MHC class II; major histocompatibility complex; autoimmune disease;

KW inflammatory disease; binding; rheumatoid arthritis; antiinflammatory;

KW antiarthritic; multiple sclerosis.

XX

OS Synthetic.

XX

PN WO200005249-A2.

XX

PD 03-FEB-2000.

XX

PF 22-JUL-1999; 99WO-US016617.

XX

PR 23-JUL-1998; 98US-0093859P.

XX

PR 09-MAR-1999; 99US-0123675P.

XX

PA (HARD ) HARVARD COLLEGE.

XX

PI Strominger JL, Fridkis-Hareli M;

XX

DR WPI; 2000-205374/18.

XX

PT New synthetic peptide, useful for treating autoimmune disease, e.g.

PT rheumatoid arthritis.

XX

PS Claim 24; Page 41; 57pp; English.

XX

CC The present invention describes synthetic peptides having an amino acid

CC sequence comprising at least 3 residues selected from the group of amino

CC acids consisting of aromatic acids, negatively charged amino acids,

CC positively charged amino acids, and aliphatic amino acids, the synthetic

CC peptides being at least 7 amino acid residues in length and capable of

CC binding to a major histocompatibility complex (MHC) class II protein

CC associated with an autoimmune disease. The synthetic peptides have anti-

CC inflammatory and anti-arthritis activities. They are used to treat

CC inflammatory and demyelinating autoimmune diseases, especially rheumatoid

CC arthritis and multiple sclerosis. The peptides are specific for

CC particular MHC Class II alleles. Purified, short and synthetic peptides

CC should have fewer side effects than mixtures of random peptides; may

CC include many repeats of the active sequence and/or contain amino acid

CC analogues that improve stability (or other desired features). AAY82021 to

CC AAY82044 represent specifically claimed peptide sequences which can be

CC used as part of the synthetic peptides of the present invention; AAY82045

CC to AAY82063 represent specifically claimed examples of the synthetic

CC peptides from the present invention; and AAY82064 to AAY82080 represent

CC other peptides used in the exemplification of the present invention

XX

SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 3; Length 15;

Best Local Similarity 91.7%; Pred. No. 1.6;

Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13

Db 4 AYASAAAAAAA 15

## RESULT 19

AAY82077

ID AAY82077 standard; peptide; 15 AA.

XX

AC AAY82077;

XX

DT 01-JUN-2000 (first entry)

XX

DE MHC class II protein binding peptide SEQ ID NO:27.

XX

KW MHC class II; major histocompatibility complex; autoimmune disease;

KW inflammatory disease; binding; rheumatoid arthritis; antiinflammatory;

KW antiarthritic; multiple sclerosis.

XX

OS Synthetic.

XX

PN WO200005249-A2.

XX

PD 03-FEB-2000.

XX

PF 22-JUL-1999; 99WO-US016617.

XX

PR 23-JUL-1998; 98US-0093859P.

XX

PR 09-MAR-1999; 99US-0123675P.

XX

PA (HARD ) HARVARD COLLEGE.

XX

PI Strominger JL, Fridkis-Hareli M;

XX

DR WPI; 2000-205374/18.

XX

PT New synthetic peptide, useful for treating autoimmune disease, e.g.

PT rheumatoid arthritis.

XX

PS Example 8; Page 32; 57pp; English.

XX

CC The present invention describes synthetic peptides having an amino acid

CC sequence comprising at least 3 residues selected from the group of amino

CC acids consisting of aromatic acids, negatively charged amino acids,

CC positively charged amino acids, and aliphatic amino acids, the synthetic

CC peptides being at least 7 amino acid residues in length and capable of

CC binding to a major histocompatibility complex (MHC) class II protein

CC associated with an autoimmune disease. The synthetic peptides have anti-

CC inflammatory and anti-arthritis activities. They are used to treat

CC inflammatory and demyelinating autoimmune diseases, especially rheumatoid

CC arthritis and multiple sclerosis. The peptides are specific for  
 CC particular MHC Class II alleles. Purified, short and synthetic peptides  
 CC should have fewer side effects than mixtures of random peptides; may  
 CC include many repeats of the active sequence and/or contain amino acid  
 CC analogues that improve stability (or other desired features). AAY82021 to  
 CC AAY82044 represent specifically claimed peptide sequences which can be  
 CC used as part of the synthetic peptides of the present invention; AAY82045  
 CC to AAY82063 represent specifically claimed examples of the synthetic  
 CC peptides from the present invention; and AAY82064 to AAY82080 represent  
 CC other peptides used in the exemplification of the present invention  
 XX  
 SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 3; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
 |:|||||  
 Db 3 AYASAAAAAAA 14

RESULT 20  
 AAG63197  
 ID AAG63197 standard; peptide; 15 AA.  
 AC AAG63197;  
 XX  
 XX  
 DT 01-OCT-2001 (first entry)  
 XX

DE Peptide which binds to peptide binding groove of HLA-DR molecules.

XX Neuronal degeneration; nerve regeneration; injury; disease; T cell;  
 KW copolymer 1; Cop 1; spinal cord injury; blunt trauma; trauma;  
 KW haemorrhagic stroke; ischaemic stroke; diabetic neuropathy; Bell's palsy;  
 KW senile dementia; Alzheimer's disease; Parkinson's disease; glaucoma;  
 KW facial nerve palsy; Huntington's chorea; amyotrophic lateral sclerosis;  
 KW status epilepticus; non-arteritic optical neuropathy; vitamin deficiency;  
 KW HLA-DR molecule.

XX Synthetic.

XX WO200152878-A2.

XX 26-JUL-2001.

XX 22-JAN-2001; 2001WO-US002117.

XX 20-JAN-2000; 2000US-00487793.

PR 07-JUN-2000; 2000US-0209799P.

PR 20-JUL-2000; 2000US-00620216.

XX (YEDA ) YEDA RES & DEV CO LTD.

PA (MCIN/) MCINNIS P A.

XX Eisenbach-Schwartz M, Cohen IR, Sela M, Yoles E, Kipnis J;

XX WPI; 2001-476094/51.

XX Promoting nerve regeneration or preventing, inhibiting or treating  
 PT neuronal degeneration caused by injury or disease comprises administering  
 PT T cells activated by copolymer 1 (Cop 1) or a Cop 1-related peptide or  
 PT polypeptide.

XX Disclosure; Page 32; 105pp; English.

XX The specification describes a method for preventing or inhibiting  
 CC neuronal degeneration, or for promoting nerve regeneration, or treating  
 CC neuronal degeneration caused by injury or disease. The method comprises  
 CC administering activated T cells which have been activated by copolymer 1  
 CC (Cop 1) or a Cop 1-related peptide or polypeptide; or Cop 1 or a Cop 1-  
 CC related peptide or polypeptide. The method is used for preventing or  
 CC inhibiting neuronal degeneration, or for promoting nerve regeneration, or

CC treating neuronal degeneration caused by injury or disease comprising  
 CC spinal cord injury, blunt trauma, penetrating trauma, haemorrhagic  
 CC stroke, ischaemic stroke, diabetic neuropathy, senile dementia,  
 CC Alzheimer's disease, Parkinson's disease, facial nerve (Bell's) palsy,  
 CC glaucoma, Huntington's chorea, amyotrophic lateral sclerosis, status  
 CC epilepticus, non-arteritic optical neuropathy, or vitamin deficiency. Cop  
 CC 1 or the Cop 1-related peptide or polypeptide is useful in the  
 CC preparation of a medicament for preventing or inhibiting neuronal  
 CC degeneration, or for promoting nerve regeneration. AAG63175-AAG63206  
 CC represent peptides which bind to the peptide binding groove of HLA-DR  
 CC molecules. The peptides are homologous to the binding motifs of Cop 1,  
 CC and have similar activities to Cop 1. As such, they may be used in the  
 CC method of the invention  
 XX  
 SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 4; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
 |:|||||  
 Db 3 AYASAAAAAAA 14

RESULT 21  
 AAG63203  
 ID AAG63203 standard; peptide; 15 AA.  
 XX  
 AC AAG63203;  
 XX

DT 01-OCT-2001 (first entry)

XX Peptide which binds to peptide binding groove of HLA-DR molecules.

XX Neuronal degeneration; nerve regeneration; injury; disease; T cell;  
 KW copolymer 1; Cop 1; spinal cord injury; blunt trauma; trauma;  
 KW haemorrhagic stroke; ischaemic stroke; diabetic neuropathy; Bell's palsy;  
 KW senile dementia; Alzheimer's disease; Parkinson's disease; glaucoma;  
 KW facial nerve palsy; Huntington's chorea; amyotrophic lateral sclerosis;  
 KW status epilepticus; non-arteritic optical neuropathy; vitamin deficiency;  
 KW HLA-DR molecule.

XX Synthetic.

XX WO200152878-A2.

XX 26-JUL-2001.

XX 22-JAN-2001; 2001WO-US002117.

XX 20-JAN-2000; 2000US-00487793.

PR 07-JUN-2000; 2000US-0209799P.

PR 20-JUL-2000; 2000US-00620216.

XX (YEDA ) YEDA RES & DEV CO LTD.

PA (MCIN/) MCINNIS P A.

XX Eisenbach-Schwartz M, Cohen IR, Sela M, Yoles E, Kipnis J;

XX WPI; 2001-476094/51.

XX Promoting nerve regeneration or preventing, inhibiting or treating  
 PT neuronal degeneration caused by injury or disease comprises administering  
 PT T cells activated by copolymer 1 (Cop 1) or a Cop 1-related peptide or  
 PT polypeptide.

XX Disclosure; Page 32; 105pp; English.

XX The specification describes a method for preventing or inhibiting  
 CC neuronal degeneration, or for promoting nerve regeneration, or treating  
 CC neuronal degeneration caused by injury or disease. The method comprises  
 CC administering activated T cells which have been activated by copolymer 1  
 CC

CC (Cop 1) or a Cop 1-related peptide or polypeptide; or Cop 1 or a Cop 1-  
CC related peptide or polypeptide. The method is used for preventing or  
CC inhibiting neuronal degeneration, or for promoting nerve regeneration, or  
CC treating neuronal degeneration caused by injury or disease comprising  
CC spinal cord injury, blunt trauma, penetrating trauma, haemorrhagic  
CC stroke, ischaemic stroke, diabetic neuropathy, senile dementia,  
CC Alzheimer's disease, Parkinson's disease, facial nerve (Bell's) palsy,  
CC glioma, Huntington's chorea, amyotrophic lateral sclerosis, status  
CC epilepticus, non-artritic optical neuropathy, or vitamin deficiency. Cop  
CC 1 or the Cop 1-related peptide or polypeptide is useful in the  
CC preparation of a medicament for preventing or inhibiting neuronal  
CC degeneration, or for promoting nerve regeneration. AAG63175-AAG63206  
CC represent peptides which bind to the peptide binding groove of HLA-DR  
CC molecules. The peptides are homologous to the binding motifs of Cop 1,  
CC and have similar activities to Cop 1. As such, they may be used in the  
CC method of the invention  
XX  
XX  
SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 4; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 AFASAAAAAAA 13  
Db 4 AYASAAAAAAA 15  
|:|||||

RESULT 22  
AAG63194  
ID AAG63194 standard; peptide; 15 AA.

AC AAG63194;

DT 01-OCT-2001 (first entry)

DE Peptide which binds to peptide binding groove of HLA-DR molecules.

XX Neuronal degeneration; nerve regeneration; injury; disease; T cell;  
XX copolymer 1; Cop 1; spinal cord injury; blunt trauma; trauma;  
KW haemorrhagic stroke; ischaemic stroke; diabetic neuropathy; Bell's palsy;  
KW senile dementia; Alzheimer's disease; Parkinson's disease; glioma;  
KW facial nerve palsy; Huntington's chorea; amyotrophic lateral sclerosis;  
KW status epilepticus; non-artritic optical neuropathy; vitamin deficiency;  
KW HLA-DR molecule.

XX Synthetic.

XX WO200152878-A2.

XX 26-JUL-2001.

XX 22-JAN-2001; 2001WO-US002117.

XX 20-JAN-2000; 2000US-00487793.

XX 07-JUN-2000; 2000US-02097992.

XX 20-JUL-2000; 2000US-00620216.

XX (YEDA) YEDA RES & DEV CO LTD.

XX (MCIN/) MCINNIS P A.

XX Eisenbach-Schwartz M, Cohen IR, Sela M, Yoles E, Kipnis J;

XX WPI; 2001-476094/51.

XX Promoting nerve regeneration or preventing, inhibiting or treating  
PT neuronal degeneration caused by injury or disease comprising administering  
PT T cells activated by copolymer 1 (Cop 1) or a Cop 1-related peptide or  
PT polypeptide.

XX Disclosure; Page 32; 105pp; English.

XX The specification describes a method for preventing or inhibiting

CC neuronal degeneration, or for promoting nerve regeneration, or treating  
CC neuronal degeneration caused by injury or disease. The method comprises  
CC administering activated T cells which have been activated by copolymer 1  
CC (Cop 1) or a Cop 1-related peptide or polypeptide; or Cop 1 or a Cop 1-  
CC related peptide or polypeptide. The method is used for preventing or  
CC inhibiting neuronal degeneration, or for promoting nerve regeneration, or  
CC treating neuronal degeneration caused by injury or disease comprising  
CC spinal cord injury, blunt trauma, penetrating trauma, haemorrhagic  
CC stroke, ischaemic stroke, diabetic neuropathy, senile dementia,  
CC Alzheimer's disease, Parkinson's disease, facial nerve (Bell's) palsy,  
CC glioma, Huntington's chorea, amyotrophic lateral sclerosis, status  
CC epilepticus, non-artritic optical neuropathy, or vitamin deficiency. Cop  
CC 1 or the Cop 1-related peptide or polypeptide is useful in the  
CC preparation of a medicament for preventing or inhibiting neuronal  
CC degeneration, or for promoting nerve regeneration. AAG63175-AAG63206  
CC represent peptides which bind to the peptide binding groove of HLA-DR  
CC molecules. The peptides are homologous to the binding motifs of Cop 1,  
CC and have similar activities to Cop 1. As such, they may be used in the  
CC method of the invention  
XX  
XX  
SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 4; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 AFASAAAAAAA 13  
Db 3 AYASAAAAAAA 14  
|:|||||

RESULT 23

ABP52216

ID ABP52216 standard; peptide; 15 AA.

AC ABP52216;

DT 16-OCT-2002 (first entry)

DE HLA-DR2 molecule binding peptide SEQ ID NO:10.

XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;  
XX HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;  
KW immune response; anti-inflammatory; neuroprotective; proliferation;  
KW MHC class II protein inhibitor; demyelinating disease; inhibition;  
KW post-viral encephalomyelitis; post-vaccine demyelinating condition;  
KW anti-tumour necrosis factor agent.

XX Homo sapiens.

OS Synthetic.

XX WO200259143-A2.

XX 01-AUG-2002.

XX 24-JAN-2002; 2002WO-US002071.

XX 24-JAN-2001; 2001US-0263569P.

XX (HARD ) HARVARD COLLEGE.

XX Strominger JL, Fridkis-Hareli M;

XX WPI; 2002-608439/65.

XX New compositions comprising synthetic peptides in complex with a major  
PT histocompatibility complex class II HLA-DR2 protein, useful for treating a  
PT demyelinating disease, e.g. multiple sclerosis, or post-viral  
PT encephalomyelitis.

XX Example 1; Page 30; 54pp; English.

XX The present invention describes compositions (I) comprising a peptide

CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine  
 CC residues. The complex of the peptide with a major histocompatibility  
 CC complex (MHC) class II HLA-DR2 protein is involved in modulating an  
 CC immune response. (I) has antiinflammatory and neuroprotective activities,  
 CC and can be used as a MHC class II protein inhibitor. The compositions  
 CC comprising the peptides are useful for treating demyelinating diseases  
 CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine  
 CC demyelinating condition, and a side effect of administering an anti-  
 CC tumour necrosis factor agents. The peptide further inhibits proliferation  
 CC of autoantigen-specific HLA-DR2-restricted T cell clones. ABP52207 to  
 CC ABP52305 represent peptides used in the exemplification of the present  
 CC invention  
 XX  
 XX Sequence 15 AA;

Query Match 87.0%; Score 47; DB 5; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
 |:|||||  
 Db 4 AYASAAAAAAA 15

RESULT 24  
 ABP52286  
 ID ABP52286 standard; peptide; 15 AA.

XX AC ABP52286;

XX DT 16-OCT-2002 (first entry)

XX DE HLA-DR2 molecule binding peptide SEQ ID NO:80.

XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;  
 KW HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;  
 KW immune response; antiinflammatory; neuroprotective; proliferation;  
 KW MHC class II protein inhibitor; demyelinating disease; inhibition;  
 KW post-viral encephalomyelitis; post-vaccine demyelinating condition;  
 KW anti-tumour necrosis factor agent.

XX OS Homo sapiens.  
 OS Synthetic.

XX PN WO200259143-A2.

XX PD 01-AUG-2002.

XX PF 24-JAN-2002; 2002WO-US002071.

XX PR 24-JAN-2001; 2001US-0263569P.

XX PA (HARD ) HARVARD COLLEGE.

XX PI Strominger JL, Fridkis-Hareli M;

XX WPI; 2002-608439/65.

XX New compositions comprising synthetic peptides in complex with a major  
 PT histocompatibility complex class II HLA-DR2 protein, useful for treating a  
 PT demyelinating disease, e.g. multiple sclerosis, or post-viral  
 PT encephalomyelitis.

XX Example 1; Page 34; 54pp; English.

XX The present invention describes compositions (I) comprising a peptide  
 CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine  
 CC residues. The complex of the peptide with a major histocompatibility  
 CC complex (MHC) class II HLA-DR2 protein is involved in modulating an  
 CC immune response. (I) has antiinflammatory and neuroprotective activities,  
 CC and can be used as a MHC class II protein inhibitor. The compositions  
 CC comprising the peptides are useful for treating demyelinating diseases  
 CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine

CC demyelinating condition, and a side effect of administering an anti-  
 CC tumour necrosis factor agents. The peptide further inhibits proliferation  
 CC of autoantigen-specific HLA-DR2-restricted T cell clones. ABP52207 to  
 CC ABP52305 represent peptides used in the exemplification of the present  
 CC invention  
 XX  
 XX Sequence 15 AA;

Query Match 87.0%; Score 47; DB 5; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
 |:|||||  
 Db 4 AYASAAAAAAA 15

RESULT 25  
 ABP52212  
 ID ABP52212 standard; peptide; 15 AA.

XX AC ABP52212;

XX DT 16-OCT-2002 (first entry)

XX DE HLA-DR2 molecule binding peptide SEQ ID NO:6.

XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;  
 KW HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;  
 KW immune response; antiinflammatory; neuroprotective; proliferation;  
 KW MHC class II protein inhibitor; demyelinating disease; inhibition;  
 KW post-viral encephalomyelitis; post-vaccine demyelinating condition;  
 KW anti-tumour necrosis factor agent.

XX OS Homo sapiens.  
 OS Synthetic.

XX PN WO200259143-A2.

XX PD 01-AUG-2002.

XX PF 24-JAN-2002; 2002WO-US002071.

XX PR 24-JAN-2001; 2001US-0263569P.

XX PA (HARD ) HARVARD COLLEGE.

XX PI Strominger JL, Fridkis-Hareli M;

XX WPI; 2002-608439/65.

XX New compositions comprising synthetic peptides in complex with a major  
 PT histocompatibility complex class II HLA-DR2 protein, useful for treating a  
 PT demyelinating disease, e.g. multiple sclerosis, or post-viral  
 PT encephalomyelitis.

XX Example 1; Page 30; 54pp; English.

XX The present invention describes compositions (I) comprising a peptide  
 CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine  
 CC residues. The complex of the peptide with a major histocompatibility  
 CC complex (MHC) class II HLA-DR2 protein is involved in modulating an  
 CC immune response. (I) has antiinflammatory and neuroprotective activities,  
 CC and can be used as a MHC class II protein inhibitor. The compositions  
 CC comprising the peptides are useful for treating demyelinating diseases  
 CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine  
 CC demyelinating condition, and a side effect of administering an anti-  
 CC tumour necrosis factor agents. The peptide further inhibits proliferation  
 CC of autoantigen-specific HLA-DR2-restricted T cell clones. ABP52207 to  
 CC ABP52305 represent peptides used in the exemplification of the present  
 CC invention

XX Sequence 15 AA;



Query Match 87.0%; Score 47; DB 5; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
|:|||||  
DB 4 AYASAAAAAAA 15

RESULT 26  
ABP52214  
ID ABP52214 standard; peptide; 15 AA.  
XX  
AC ABP52214;  
XX  
DT 16-OCT-2002 (first entry)  
XX  
DE HLA-DR2 molecule binding peptide SEQ ID NO:8.

XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;  
KW HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;  
KW immune response; antiinflammatory; neuroprotective; proliferation;  
KW MHC class II protein inhibitor; demyelinating disease; inhibition;  
KW post-viral encephalomyelitis; post-vaccine demyelinating condition;  
KW anti-tumour necrosis factor agent.

XX Homo sapiens.  
OS Synthetic.  
OS  
PN WO200259143-A2.  
XX  
XX 01-AUG-2002.  
XX  
XX 24-JAN-2002; 2002WO-US002071.  
XX  
XX 24-JAN-2001; 2001US-0263569P.  
XX (HARD ) HARVARD COLLEGE.

XX Strominger JL, Fridkis-Hareli M;  
XX WPI; 2002-608439/65.  
XX

XX New compositions comprising synthetic peptides in complex with a major  
PT histocompatibility complex class II HLA-DR2 protein, useful for treating a  
PT demyelinating disease, e.g. multiple sclerosis, or post-viral  
PT encephalomyelitis.

XX Example 1; Page 30; 54pp; English.

XX The present invention describes compositions (I) comprising a peptide  
CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine  
CC residues. The complex of the peptide with a major histocompatibility  
CC complex (MHC) class II HLA-DR2 protein is involved in modulating an  
CC immune response. (I) has antiinflammatory and neuroprotective activities,  
CC and can be used as a MHC class II protein inhibitor. The compositions  
CC comprising the peptides are useful for treating demyelinating diseases  
CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine  
CC demyelinating condition, and a side effect of administering an anti-  
CC tumour necrosis factor agents. The peptide further inhibits proliferation  
CC of autoantigen-specific HLA-DR2-restricted T cell clones. ABP52207 to  
CC ABP52305 represent peptides used in the exemplification of the present  
CC invention

XX Sequence 15 AA;

Query Match 87.0%; Score 47; DB 5; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
|:|||||

DB 4 AYASAAAAAAA 15

RESULT 27  
ABP52210  
ID ABP52210 standard; peptide; 15 AA.  
XX  
AC ABP52210;  
XX

XX 16-OCT-2002 (first entry)

XX HLA-DR2 molecule binding peptide SEQ ID NO:4.

XX Human leukocyte antigen; HLA; major histocompatibility complex; MHC;  
KW HLA-DR2; MHC class II DR-2 molecule; demyelinating; multiple sclerosis;  
KW immune response; antiinflammatory; neuroprotective; proliferation;  
KW MHC class II protein inhibitor; demyelinating disease; inhibition;  
KW post-viral encephalomyelitis; post-vaccine demyelinating condition;  
KW anti-tumour necrosis factor agent.

XX Homo sapiens.  
OS Synthetic.  
OS  
PN WO200259143-A2.  
XX  
XX 01-AUG-2002.

XX 24-JAN-2002; 2002WO-US002071.

XX 24-JAN-2001; 2001US-0263569P.

XX (HARD ) HARVARD COLLEGE.

XX Strominger JL, Fridkis-Hareli M;

XX WPI; 2002-608439/65.

XX New compositions comprising synthetic peptides in complex with a major  
PT histocompatibility complex class II HLA-DR2 protein, useful for treating a  
PT demyelinating disease, e.g. multiple sclerosis, or post-viral  
PT encephalomyelitis.

XX Example 1; Page 30; 54pp; English.

XX The present invention describes compositions (I) comprising a peptide  
CC with an amino acid sequence with tyrosine (Y), lysine (K), and/or valine  
CC residues. The complex of the peptide with a major histocompatibility  
CC complex (MHC) class II HLA-DR2 protein is involved in modulating an  
CC immune response. (I) has antiinflammatory and neuroprotective activities,  
CC and can be used as a MHC class II protein inhibitor. The compositions  
CC comprising the peptides are useful for treating demyelinating diseases  
CC such as multiple sclerosis, post-viral encephalomyelitis, a post-vaccine  
CC demyelinating condition, and a side effect of administering an anti-  
CC tumour necrosis factor agents. The peptide further inhibits proliferation  
CC of autoantigen-specific HLA-DR2-restricted T cell clones. ABP52207 to  
CC ABP52305 represent peptides used in the exemplification of the present  
CC invention

XX Sequence 15 AA;

Query Match 87.0%; Score 47; DB 5; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
|:|||||  
DB 4 AYASAAAAAAA 15

RESULT 28

ADJ57538

ID ADJ57538 standard; peptide; 15 AA.

XX



XX Vaccine useful in the treatment of motor neuron disease e.g. amyotrophic  
PT lateral sclerosis and progressive muscular atrophy, comprises copolymer-1  
PT or related peptide or polypeptide, or poly-Glu,Tyr.  
XX  
PS Disclosure; SEQ ID NO 23; 49pp; English.  
XX  
XX The invention relates to a vaccine that comprises an active agent from  
CC Copolymer 1 (Cop-1), Cop-1 related peptide, Cop-1 related polypeptide or  
CC poly-Glu,Tyr. The vaccine comprises the active agent without adjuvant or  
CC is emulsified in an adjuvant. The vaccine is useful for reducing disease  
CC progression and/or protection of motor nerve degeneration and/or  
CC protection from glutamate toxicity in motor neuron disease e.g.  
CC amyotrophic lateral sclerosis, primary lateral sclerosis, progressive  
CC muscular atrophy, progressive bulbar palsy (PBP or bulbar onset). It may  
CC be used in combination with another drug e.g. Riluzole for treating motor  
CC neuron disease. The non-pathogenic synthetic copolymers protects  
CC autoimmunity and thus reduces further injury-induced damage and protects  
CC central nervous system cells from glutamate toxicity; protects retinal  
CC ganglion cells from glutamate toxicity. The present sequence represents a  
CC Cop-1 related-peptide.  
XX  
SQ Sequence 15 AA;  
  
Query Match 87.0%; Score 47; DB 7; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 AFRAAAAAAAAA 13  
|:|||||  
DB 3 AYAAAAAAAAA 14  
  
RESULT 31  
ADQS9600  
ID ADQS9600 standard; peptide; 15 AA.  
XX  
AC ADQS9600;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
XX Copolymer 1-related peptide SEQ ID NO:23.  
XX  
XX eye-drop vaccine; eye drop; vaccine; immunisation; Copolymer 1;  
KW Copolymer 1-related peptide; ophthalmological; neuroprotective;  
KW nootropic; antiparkinsonian; anticonvulsant; analgesic; antidiabetic;  
KW hepatotropic; respiratory; osteopathic; neuronal degeneration;  
KW central nervous system; CNS; peripheral nervous system; PNS;  
KW glutamate toxicity; spinal cord injury; blunt trauma;  
KW haemorrhagic stroke; ischaemic stroke; senile dementia;  
KW Alzheimer's disease; Parkinson's disease; motor neuron disease;  
KW amyotrophic lateral sclerosis; prion disease; Creutzfeldt-Jakob disease;  
KW vitamin deficiency; epilepsy; amnesia; anxiety; autoimmune disease;  
KW multiple sclerosis; peripheral neuropathy; retinal disorder; glaucoma.  
XX  
OS Synthetic.  
XX  
XX WO2004060265-A2.  
XX  
XX 22-JUL-2004.  
XX  
XX 06-JAN-2004; 2004WO-IL000006.  
XX  
XX 07-JAN-2003; 2003US-0438310P.  
XX  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX  
XX Eisenbach-Schwartz M, Bakalash S, Fulga V;  
XX  
XX WPI; 2004-543774/52.  
XX  
XX Eye-drop vaccine for the therapeutic immunization of a mammal against  
PT glaucoma and neuronal degeneration caused by a disease in the central or

peripheral nervous system (e.g. Alzheimer's disease), comprises Copolymer  
1.  
XX  
PS Disclosure; SEQ ID NO 23; 49pp; English.  
XX  
XX The present invention describes an eye-drop vaccine (I) for the  
CC therapeutic immunisation of a mammal. (I) comprises an active agent  
CC selected from Copolymer 1, a Copolymer 1-related peptide, and a Copolymer  
CC 1-related polypeptide. (I) has ophthalmological, neuroprotective,  
CC nootropic, antiparkinsonian, anticonvulsant, analgesic, antidiabetic,  
CC hepatotropic, respiratory and osteopathic activities. The eye-drop  
CC vaccine (I) can be used for treating or preventing neuronal degeneration  
CC caused by an injury, disease, disorder or condition in the central  
CC nervous system (CNS) or peripheral nervous system (PNS), for preventing  
CC or inhibiting neuronal secondary degeneration which may otherwise follow  
CC a primary injury in the CNS, for promoting nerve regeneration in the CNS  
CC or in the PNS after an injury, disease, disorder or condition or for  
CC protecting CNS and PNS cells from glutamate toxicity, where the injury is  
CC spinal cord injury, blunt trauma, penetrating trauma, brain coup or  
CC contrecoup, haemorrhagic stroke or ischaemic stroke, and where the  
CC disease is a senile dementia including Alzheimer's disease, a  
CC Parkinsonian syndrome including Parkinson's disease, facial nerve  
CC (Bell's) palsy, Huntington's chorea, a motor neuron disease including  
CC amyotrophic lateral sclerosis, a prion disease including Creutzfeldt-  
CC Jakob disease, Alper's disease, Batten disease, Cockayne syndrome, Lewy  
CC body disease, status epilepticus, carpal tunnel syndrome, intervertebral  
CC disc herniation, vitamin deficiency such as vitamin B deficiency,  
CC epilepsy, amnesia, anxiety, hyperalgesia, psychosis, seizures, oxidative  
CC stress, opiate tolerance and dependence, an autoimmune disease (e.g.  
CC multiple sclerosis), or a peripheral neuropathy associated with a disease  
CC such as amyloid polynuropathy, diabetic neuropathy, uremic neuropathy,  
CC porphyric polynuropathy, hypoglycaemia, Sjogren-Larsson syndrome, acute  
CC sensory neuropathy, chronic ataxic neuropathy, biliary cirrhosis, primary  
CC amyloidosis, obstructive lung diseases, acromegaly, malabsorption  
CC syndromes, polycythemia vera, IGA and IGG gammopathies, complications of  
CC various drugs such as nitrofurantoin, metronidazole, isoniazid and toxins  
CC such as alcohol or organophosphates, Charcot-Marie-Tooth disease, ataxia  
CC telangiectasia, Friedreich's ataxia, adrenomyeloneuropathy, giant axonal  
CC neuropathy, Refsum's disease, Fabry's disease, lipoproteinemia, non-  
CC arteritic optic neuropathy, age-related macular degeneration, a retinal  
CC disorder such as retinal degeneration, or a disease associated with  
CC abnormally elevated intraocular pressure such as glaucoma. The present  
CC sequence represents a copolymer 1 related peptide, which is given in the  
CC exemplification of the present invention.  
XX  
SQ Sequence 15 AA;  
  
Query Match 87.0%; Score 47; DB 8; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 AFRAAAAAAAAA 13  
|:|||||  
DB 3 AYAAAAAAAAA 14  
  
RESULT 32  
ADQS9606  
ID ADQS9606 standard; peptide; 15 AA.  
XX  
AC ADQS9606;  
XX  
DT 07-OCT-2004 (first entry)  
XX  
XX Copolymer 1-related peptide SEQ ID NO:29.  
XX  
XX eye-drop vaccine; eye drop; vaccine; immunisation; Copolymer 1;  
KW Copolymer 1-related peptide; ophthalmological; neuroprotective;  
KW nootropic; antiparkinsonian; anticonvulsant; analgesic; antidiabetic;  
KW hepatotropic; respiratory; osteopathic; neuronal degeneration;  
KW central nervous system; CNS; peripheral nervous system; PNS;  
KW glutamate toxicity; spinal cord injury; blunt trauma;  
KW haemorrhagic stroke; ischaemic stroke; senile dementia;  
KW

KW Alzheimer's disease; parkinson's disease; motor neuron disease;  
 KW amyotrophic lateral sclerosis; prion disease; Creutzfeldt-Jakob disease;  
 KW vitamin deficiency; epilepsy; amnesia; anxiety; autoimmune disease;  
 XX multiple sclerosis; peripheral neuropathy; retinal disorder; glaucoma.  
 OS Synthetic.  
 XX WO2004060265-A2.  
 XX 22-JUL-2004.  
 PD 06-JAN-2004; 2004WO-IL000006.  
 XX 07-JAN-2003; 2003US-0438310P.  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 XX Eisenbach-Schwartz M, Bakalash S, Fulga V;  
 XX WPI; 2004-543774/52.  
 XX Eye-drop vaccine for the therapeutic immunization of a mammal against  
 PT glaucoma and neuronal degeneration caused by a disease in the central or  
 PT peripheral nervous system (e.g. Alzheimer's disease), comprises Copolymer  
 PT 1.  
 XX Disclosure; SEQ ID NO 29; 49pp; English.  
 CC The present invention describes an eye-drop vaccine (I) for the  
 CC therapeutic immunisation of a mammal. (I) comprises an active agent  
 CC selected from Copolymer 1, a Copolymer 1-related peptide, and a Copolymer  
 CC 1-related polypeptide. (I) has ophthalmological, neuroprotective,  
 CC neurotropic, antiparkinsonian, anticonvulsant, analgesic, antidiabetic,  
 CC hepatotropic, respiratory and osteopathic activities. The eye-drop  
 CC vaccine (I) can be used for treating or preventing neuronal degeneration  
 CC caused by an injury, disease, disorder or condition in the central  
 CC nervous system (CNS) or peripheral nervous system (PNS), for preventing  
 CC or inhibiting neuronal secondary degeneration which may otherwise follow  
 CC a primary injury in the CNS, for promoting nerve regeneration in the CNS  
 CC or in the PNS after an injury, disease, disorder or condition or for  
 CC protecting CNS and PNS cells from glutamate toxicity, where the injury is  
 CC spinal cord injury, blunt trauma, penetrating trauma, brain coup or  
 CC contrecoup, haemorrhagic stroke or ischaemic stroke, and where the  
 CC disease is a senile dementia including Alzheimer's disease, a  
 CC Parkinsonian syndrome including Parkinson's disease, facial nerve  
 CC (Bell's) palsy, Huntington's chorea, a motor neuron disease including  
 CC amyotrophic lateral sclerosis, a prion disease including Creutzfeldt-  
 CC Jakob disease, Alper's disease, Batten disease, Cockayne syndrome, Lewy  
 CC body disease, status epilepticus, carpal tunnel syndrome, intervertebral  
 CC disc herniation, vitamin deficiency such as vitamin B deficiency,  
 CC epilepsy, amnesia, anxiety, hyperalgesia, psychosis, seizures, oxidative  
 CC stress, opiate tolerance and dependence, an autoimmune disease (e.g.  
 CC multiple sclerosis), or a peripheral neuropathy associated with a disease  
 CC such as anyloid polynuropathy, diabetic neuropathy, uremic neuropathy,  
 CC porphyric polynuropathy, hypoglycaemia, Sjogren-Larsson syndrome, acute  
 CC sensory neuropathy, chronic ataxic neuropathy, biliary cirrhosis, primary  
 CC amyloidosis, obstructive lung diseases, acromegaly, malabsorption  
 CC syndromes, polycythemia vera, IgA and IgG gammopathies, complications of  
 CC various drugs such as nitrofurantoin, metronidazole, isoniazid and toxins  
 CC such as alcohol or organophosphates, Charcot-Marie-Tooth disease, ataxia  
 CC telangiectasia, Friedreich's ataxia, adrenomyeloneuropathy, giant axonal  
 CC neuropathy, Refsum's disease, Fabry's disease, lipoproteinemia, non-  
 CC arteritic optic neuropathy, age-related macular degeneration, a retinal  
 CC disorder such as retinal degeneration, or a disease associated with  
 CC abnormally elevated intraocular pressure such as glaucoma. The present  
 CC sequence represents a copolymer 1 related peptide, which is given in the  
 CC exemplification of the present invention.  
 XX Sequence 15 AA;  
 SQ Query Match 87.0%; Score 47; DB 8; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 AFASAAAAAAA 13  
 DB 4 AYASAAAAAAA 15  
 RESULT 33  
 ADQ59597  
 ID ADQ59597 standard; peptide; 15 AA.  
 XX  
 AC ADQ59597;  
 XX  
 DT 07-OCT-2004 (first entry)  
 XX  
 DE Copolymer 1-related peptide SEQ ID NO:20.  
 XX  
 KW eye-drop vaccine; eye drop; vaccine; immunisation; Copolymer 1;  
 KW Copolymer 1-related peptide; ophthalmological; neuroprotective;  
 KW neurotropic; antiparkinsonian; anticonvulsant; analgesic; antidiabetic;  
 KW hepatotropic; respiratory; osteopathic; neuronal degeneration;  
 KW central nervous system; CNS; peripheral nervous system; PNS;  
 KW glutamate toxicity; spinal cord injury; blunt trauma;  
 KW haemorrhagic stroke; ischaemic stroke; senile dementia;  
 KW Alzheimer's disease; Parkinson's disease; motor neuron disease;  
 KW amyotrophic lateral sclerosis; prion disease; Creutzfeldt-Jakob disease;  
 KW vitamin deficiency; epilepsy; amnesia; anxiety; autoimmune disease;  
 KW multiple sclerosis; peripheral neuropathy; retinal disorder; glaucoma.  
 XX Synthetic.  
 XX WO2004060265-A2.  
 XX 22-JUL-2004.  
 XX 06-JAN-2004; 2004WO-IL000006.  
 XX 07-JAN-2003; 2003US-0438310P.  
 XX (YEDA ) YEDA RES & DEV CO LTD.  
 XX Eisenbach-Schwartz M, Bakalash S, Fulga V;  
 XX WPI; 2004-543774/52.  
 XX Eye-drop vaccine for the therapeutic immunization of a mammal against  
 PT glaucoma and neuronal degeneration caused by a disease in the central or  
 PT peripheral nervous system (e.g. Alzheimer's disease), comprises Copolymer  
 PT 1.  
 XX Disclosure; SEQ ID NO 20; 49pp; English.  
 CC The present invention describes an eye-drop vaccine (I) for the  
 CC therapeutic immunisation of a mammal. (I) comprises an active agent  
 CC selected from Copolymer 1, a Copolymer 1-related peptide, and a Copolymer  
 CC 1-related polypeptide. (I) has ophthalmological, neuroprotective,  
 CC neurotropic, antiparkinsonian, anticonvulsant, analgesic, antidiabetic,  
 CC hepatotropic, respiratory and osteopathic activities. The eye-drop  
 CC vaccine (I) can be used for treating or preventing neuronal degeneration  
 CC caused by an injury, disease, disorder or condition in the central  
 CC nervous system (CNS) or peripheral nervous system (PNS), for preventing  
 CC or inhibiting neuronal secondary degeneration which may otherwise follow  
 CC a primary injury in the CNS, for promoting nerve regeneration in the CNS  
 CC or in the PNS after an injury, disease, disorder or condition or for  
 CC protecting CNS and PNS cells from glutamate toxicity, where the injury is  
 CC spinal cord injury, blunt trauma, penetrating trauma, brain coup or  
 CC contrecoup, haemorrhagic stroke or ischaemic stroke, and where the  
 CC disease is a senile dementia including Alzheimer's disease, a  
 CC Parkinsonian syndrome including Parkinson's disease, facial nerve  
 CC (Bell's) palsy, Huntington's chorea, a motor neuron disease including  
 CC amyotrophic lateral sclerosis, a prion disease including Creutzfeldt-  
 CC Jakob disease, Alper's disease, Batten disease, Cockayne syndrome, Lewy  
 CC body disease, status epilepticus, carpal tunnel syndrome, intervertebral  
 CC disc herniation, vitamin deficiency such as vitamin B deficiency,  
 CC epilepsy, amnesia, anxiety, hyperalgesia, psychosis, seizures, oxidative  
 CC stress, opiate tolerance and dependence, an autoimmune disease (e.g.  
 CC multiple sclerosis), or a peripheral neuropathy associated with a disease  
 CC such as anyloid polynuropathy, diabetic neuropathy, uremic neuropathy,  
 CC porphyric polynuropathy, hypoglycaemia, Sjogren-Larsson syndrome, acute  
 CC sensory neuropathy, chronic ataxic neuropathy, biliary cirrhosis, primary  
 CC amyloidosis, obstructive lung diseases, acromegaly, malabsorption  
 CC syndromes, polycythemia vera, IgA and IgG gammopathies, complications of  
 CC various drugs such as nitrofurantoin, metronidazole, isoniazid and toxins  
 CC such as alcohol or organophosphates, Charcot-Marie-Tooth disease, ataxia  
 CC telangiectasia, Friedreich's ataxia, adrenomyeloneuropathy, giant axonal  
 CC neuropathy, Refsum's disease, Fabry's disease, lipoproteinemia, non-  
 CC arteritic optic neuropathy, age-related macular degeneration, a retinal  
 CC disorder such as retinal degeneration, or a disease associated with  
 CC abnormally elevated intraocular pressure such as glaucoma. The present  
 CC sequence represents a copolymer 1 related peptide, which is given in the  
 CC exemplification of the present invention.

CC epilepsy, amnesia, anxiety, hyperalgesia, psychosis, seizures, oxidative  
 CC stress, opiate tolerance and dependence, an autoimmune disease (e.g.  
 CC multiple sclerosis), or a peripheral neuropathy associated with a disease  
 CC such as amyloid polyneuropathy, diabetic neuropathy, uremic neuropathy,  
 CC porphyric polyneuropathy, hypoglycaemia, Sjogren-Larsson syndrome, acute  
 CC sensory neuropathy, chronic ataxic neuropathy, biliary cirrhosis, primary  
 CC amyloidosis, obstructive lung diseases, acromegaly, malabsorption  
 CC syndromes, polycythemia vera, IGA and IgG gammopathies, complications of  
 CC various drugs such as nitrofurantoin, metronidazole, isoniazid and toxins  
 CC such as alcohol or organophosphates, Charcot-Marie-Tooth disease, ataxia  
 CC telangiectasia, Friedreich's ataxia, adrenomyeloneuropathy, giant axonal  
 CC neuropathy, Refsum's disease, Fabry's disease, lipoproteinemia, non-  
 CC arteritic optic neuropathy, age-related macular degeneration, a retinal  
 CC disorder such as retinal degeneration, or a disease associated with  
 CC abnormally elevated intraocular pressure such as glaucoma. The present  
 CC sequence represents a copolymer 1 related peptide, which is given in the  
 CC exemplification of the present invention.

XX Sequence 15 AA;

Query Match 87.0%; Score 47; DB 8; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AFAAAAAAAAAA 13  
 Db 3 AYAAAAAAAAAA 14

RESULT 34  
 ADQ81224  
 ID ADQ81224 standard; peptide; 15 AA.

XX AC ADQ81224;

XX DT 21-OCT-2004 (first entry)

XX DE Copolymer 1-related peptide SEQ ID NO:20.

XX KW pharmaceutical composition; copolymer 1; copolymer 1-related polypeptide;  
 KW copolymer 1-related peptide; inflammatory bowel disease;  
 KW antiinflammatory; antiulcer; gastrointestinal; regulatory T cell inducer;  
 KW Crohn's disease; ulcerative colitis.

XX OS Synthetic.

XX PN WO2004064717-A2.

XX PD 05-AUG-2004.

XX PF 20-JAN-2004; 2004WO-IL000054.

XX PR 21-JAN-2003; 2003US-0441136P.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PI Aharoni R, Arnon R, Kayhan B;

XX DR WPI; 2004-571593/55.

XX PT Composition useful for the treatment of a patient suffering from an  
 PT inflammatory bowel disease comprises an active agent selected from  
 PT copolymer 1, a copolymer 1-related polypeptide, and a copolymer 1-related  
 PT peptide and a carrier.

XX PS Example 2; SEQ ID NO 20; 55pp; English.

XX CC The present invention describes a pharmaceutical composition comprising  
 CC an active agent selected from copolymer 1, a copolymer 1-related  
 CC polypeptide, and a copolymer 1-related peptide and a carrier. Also  
 CC described is an article of manufacture comprising packaging material and  
 CC the pharmaceutical composition (preferably containing copolymer 1)  
 CC contained within the packaging material. The packaging material includes

CC a label that indicates that the agent is used for treating an  
 CC inflammatory bowel disease. The pharmaceutical composition has  
 CC antiinflammatory, antiulcer and gastrointestinal activities, and can be  
 CC used as an inducer of regulatory T cells of the Th2 type. It can be used  
 CC for the treatment of a patient suffering from an inflammatory bowel  
 CC disease e.g. Crohn's disease and ulcerative colitis. The copolymer 1  
 CC binds promiscuously and with high affinity to various classes II MHC  
 CC molecules of mouse and human origin, and can even displace antigens from  
 CC the MHC antigen-binding groove. In this way, the presentation of other  
 CC antigens and, consequently, the persistence of inflammatory process, are  
 CC down regulated. In addition, Copolymer 1 is a potent inducer of  
 CC regulatory T cells of the Th2 type. The present sequence represents a  
 CC copolymer 1-related peptide which is used in the exemplification of the  
 CC present invention.

XX Sequence 15 AA;

Query Match 87.0%; Score 47; DB 8; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AFAAAAAAAAAA 13  
 Db 3 AYAAAAAAAAAA 14

RESULT 35  
 ADQ81233  
 ID ADQ81233 standard; peptide; 15 AA.

XX AC ADQ81233;

XX DT 21-OCT-2004 (first entry)

XX DE Copolymer 1-related peptide SEQ ID NO:29.

XX KW pharmaceutical composition; copolymer 1; copolymer 1-related polypeptide;  
 KW copolymer 1-related peptide; inflammatory bowel disease;  
 KW antiinflammatory; antiulcer; gastrointestinal; regulatory T cell inducer;  
 KW Crohn's disease; ulcerative colitis.

XX OS Synthetic.

XX PN WO2004064717-A2.

XX PD 05-AUG-2004.

XX PF 20-JAN-2004; 2004WO-IL000054.

XX PR 21-JAN-2003; 2003US-0441136P.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PI Aharoni R, Arnon R, Kayhan B;

XX DR WPI; 2004-571593/55.

XX PT Composition useful for the treatment of a patient suffering from an  
 PT inflammatory bowel disease comprises an active agent selected from  
 PT copolymer 1, a copolymer 1-related polypeptide, and a copolymer 1-related  
 PT peptide and a carrier.

XX PS Example 2; SEQ ID NO 29; 55pp; English.

XX CC The present invention describes a pharmaceutical composition comprising  
 CC an active agent selected from copolymer 1, a copolymer 1-related  
 CC polypeptide, and a copolymer 1-related peptide and a carrier. Also  
 CC described is an article of manufacture comprising packaging material and  
 CC the pharmaceutical composition (preferably containing copolymer 1)  
 CC contained within the packaging material. The packaging material includes  
 CC a label that indicates that the agent is used for treating an  
 CC inflammatory bowel disease. The pharmaceutical composition has  
 CC antiinflammatory, antiulcer and gastrointestinal activities, and can be

CC used as an inducer of regulatory T cells of the Th2 type. It can be used  
 CC for the treatment of a patient suffering from an inflammatory bowel  
 CC disease e.g. Crohn's disease and ulcerative colitis. The copolymer 1  
 CC binds promiscuously and with high affinity to various classes II MHC  
 CC molecules of mouse and human origin, and can even displace antigens from  
 CC the MHC antigen-binding groove. In this way, the presentation of other  
 CC antigens and consequently, the persistence of inflammatory process, are  
 CC down regulated. In addition, Copolymer 1 is a potent inducer of  
 CC regulatory T cells of the Th2 type. The present sequence represents a  
 CC copolymer 1-related peptide which is used in the exemplification of the  
 CC present invention.

XX SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 8; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AFMAAAAAAAAA 13  
 |:|||||  
 Db 4 AYAAAAAAAAAAAA 15

# RESULT 36

ID ADQ81227 standard; peptide; 15 AA.

XX AC ADQ81227;

DT 21-OCT-2004 (first entry)

XX Copolymer 1-related peptide SEQ ID NO:23.

DE pharmaceutical composition; copolymer 1; copolymer 1-related polypeptide;  
 XX copolymer 1-related peptide; inflammatory bowel disease;  
 KW antiinflammatory; antiulcer; gastrointestinal; regulatory T cell inducer;  
 KW Crohn's disease; ulcerative colitis.

XX OS Synthetic.

XX PN WO2004064717-A2.

XX PD 05-AUG-2004.

XX PF 20-JAN-2004; 2004WO-IL000054.

XX PR 21-JAN-2003; 2003US-0441136P.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PI Aharoni R, Arnon R, Kayhan B;

XX DR WPI; 2004-571593/55.

XX Composition useful for the treatment of a patient suffering from an  
 PT inflammatory bowel disease comprises an active agent selected from  
 PT copolymer 1, a copolymer 1-related polypeptide, and a copolymer 1-related  
 PT peptide and a carrier.

XX Example 2; SEQ ID NO 23; 55pp; English.

XX The present invention describes a pharmaceutical composition comprising  
 CC an active agent selected from copolymer 1, a copolymer 1-related  
 CC polypeptide, and a copolymer 1-related peptide and a carrier. Also  
 CC described is an article of manufacture comprising packaging material and  
 CC the pharmaceutical composition (preferably containing copolymer 1)  
 CC contained within the packaging material. The packaging material includes  
 CC a label that indicates that the agent is used for treating an  
 CC inflammatory bowel disease. The pharmaceutical composition has  
 CC antiinflammatory, antiulcer and gastrointestinal activities, and can be  
 CC used as an inducer of regulatory T cells of the Th2 type. It can be used  
 CC for the treatment of a patient suffering from an inflammatory bowel  
 CC disease e.g. Crohn's disease and ulcerative colitis. The copolymer 1

CC binds promiscuously and with high affinity to various classes II MHC  
 CC molecules of mouse and human origin, and can even displace antigens from  
 CC the MHC antigen-binding groove. In this way, the presentation of other  
 CC antigens and consequently, the persistence of inflammatory process, are  
 CC down regulated. In addition, Copolymer 1 is a potent inducer of  
 CC regulatory T cells of the Th2 type. The present sequence represents a  
 CC copolymer 1-related peptide which is used in the exemplification of the  
 CC present invention.

XX SQ Sequence 15 AA;

Query Match 87.0%; Score 47; DB 8; Length 15;  
 Best Local Similarity 91.7%; Pred. No. 1.6;  
 Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AFMAAAAAAAAA 13  
 |:|||||

Db 3 AYAAAAAAAAAAAA 14

# RESULT 37

ADW86548

ID ADW86548 standard; peptide; 15 AA.

XX AC ADW86548;

XX DT 21-APR-2005 (first entry)

XX DE Ordered copolymer peptide used to treat graft versus host disease Seq 29.

XX KW graft rejection; immunosuppressive; immunomodulator;

XX OS Synthetic.

XX PN WO2005009333-A2.

XX PD 03-FEB-2005.

XX PF 29-JUL-2004; 2004WO-IL000695.

XX PR 31-JUL-2003; 2003US-0491236P.

XX PA (YEDA ) YEDA RES & DEV CO LTD.

XX PI (MORR-) MOR RES APPL LTD.

XX PI Aharoni R, Arnon R, Sela M, Yussim A;

XX DR WPI; 2005-152159/16.

XX Treating or preventing graft rejection associated with transplantation of  
 PT cells, tissue or organs comprises administering combination comprising  
 PT copolymer 1 or its related heteropolymer with immunosuppressive drug for.

XX Claim 3; SEQ ID NO 29; 63pp; English.

XX This invention relates to a novel method for treating or preventing graft  
 CC rejection. Specifically, it refers to administering at least one  
 CC copolymer 1 or its related heteropolymer with at least one  
 CC immunosuppressive drug. The present invention describes the copolymer as  
 CC containing at least three different amino acids each selected from at  
 CC least three of the following groups: lysine and arginine; glutamic acid  
 CC and aspartic acid; alanine, glycine and alanine; and tyrosine, tryptophan  
 CC and phenylalanine. In addition, the immunosuppressive drug is an  
 CC antiproliferative drug, inhibitor of lymphocyte activation, steroid,  
 CC purine antimetabolite, antibody or immunomodulator. Accordingly, these  
 CC immunosuppressive compositions can be used for treating or preventing  
 CC graft rejection associated with transplantation of cells, tissue or  
 CC organs (HLA-matched or mismatched) selected from hematopoietic cells,  
 CC stem cells, heart, lung, kidney, liver or skin. The copolymer 1 or its  
 CC related heteropolymer in combination with other immunosuppressive drugs  
 CC increases the effectiveness of the drugs at lower dosages, reducing toxic  
 CC side effects such that they are useful for host-versus-graft (HVG)

CC response, as well as graft-versus-host disease (GVHD). This peptide  
CC sequence is an ordered copolymer of the invention.  
XX  
SQ Sequence 15 AA;  
Query Match 87.0%; Score 47; DB 9; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AFAAAAAAAAAA 13  
|:|||||  
DB 4 AYAAAAAAAAAA 15  
|:|||||  
RESULT 38  
ADW86539  
ID ADW86539 standard; peptide; 15 AA.  
XX AC ADW86539;  
XX 21-APR-2005 (first entry)  
DE Ordered copolymer peptide used to treat graft versus host disease Seq 20.  
XX graft rejection; immunosuppressive; immunomodulator;  
KW graft versus host disease.  
XX Synthetic.  
XX WO2005009333-A2.  
XX 03-FEB-2005.  
XX 29-JUL-2004; 2004WO-IL000695.  
XX 31-JUL-2003; 2003US-0491236P.  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX (MORR-) MOR RES APPL LTD.  
PI Aharoni R, Arnon R, Sela M, Yussim A;  
XX WPI; 2005-152159/16.  
XX Treating or preventing graft rejection associated with transplantation of  
PT cells, tissue or organs comprises administering combination comprising  
PT copolymer 1 or its related heteropolymer with immunosuppressive drug for.  
XX Claim 3; SEQ ID NO 20; 63pp; English.  
XX This invention relates to a novel method for treating or preventing graft  
CC rejection. Specifically, it refers to administering at least one  
CC copolymer 1 or its related heteropolymer with at least one  
CC immunosuppressive drug. The present invention describes the copolymer as  
CC containing at least three different amino acids each selected from at  
CC least three of the following groups: lysine and arginine; glutamic acid  
CC and aspartic acid; alanine, glycine and alanine; and tyrosine, tryptophan  
CC and phenylalanine. In addition, the immunosuppressive drug is an  
CC antiproliferative drug, inhibitor of lymphocyte activation, steroid,  
CC purine antimetabolite, antibody or immunomodulator. Accordingly, these  
CC immunosuppressive compositions can be used for treating or preventing  
CC graft rejection associated with transplantation of cells, tissue or  
CC organs (HLA-matched or mismatched) selected from hematopoietic cells,  
CC stem cells, heart, lung, kidney, liver or skin. The copolymer 1 or its  
CC related heteropolymer in combination with other immunosuppressive drugs  
CC increases the effectiveness of the drugs at lower dosages, reducing toxic  
CC side effects such that they are useful for host-versus-graft (HVG)  
CC response, as well as graft-versus-host disease (GVHD). This peptide  
CC sequence is an ordered copolymer of the invention.  
XX Sequence 15 AA;  
Query Match 87.0%; Score 47; DB 9; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AFAAAAAAAAAA 13  
|:|||||  
DB 4 AYAAAAAAAAAA 15  
|:|||||

Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AFAAAAAAAAAA 13  
|:|||||  
DB 3 AYAAAAAAAAAA 14  
|:|||||  
RESULT 39  
ADW86542  
ID ADW86542 standard; peptide; 15 AA.  
XX AC ADW86542;  
XX 21-APR-2005 (first entry)  
DE Ordered copolymer peptide used to treat graft versus host disease Seq 23.  
XX graft rejection; immunosuppressive; immunomodulator;  
KW graft versus host disease.  
XX Synthetic.  
XX WO2005009333-A2.  
XX 03-FEB-2005.  
XX 29-JUL-2004; 2004WO-IL000695.  
XX 31-JUL-2003; 2003US-0491236P.  
XX (YEDA ) YEDA RES & DEV CO LTD.  
XX (MORR-) MOR RES APPL LTD.  
PI Aharoni R, Arnon R, Sela M, Yussim A;  
XX WPI; 2005-152159/16.  
XX Treating or preventing graft rejection associated with transplantation of  
PT cells, tissue or organs comprises administering combination comprising  
PT copolymer 1 or its related heteropolymer with immunosuppressive drug for.  
XX Claim 3; SEQ ID NO 23; 63pp; English.  
XX This invention relates to a novel method for treating or preventing graft  
CC rejection. Specifically, it refers to administering at least one  
CC copolymer 1 or its related heteropolymer with at least one  
CC immunosuppressive drug. The present invention describes the copolymer as  
CC containing at least three different amino acids each selected from at  
CC least three of the following groups: lysine and arginine; glutamic acid  
CC and aspartic acid; alanine, glycine and alanine; and tyrosine, tryptophan  
CC and phenylalanine. In addition, the immunosuppressive drug is an  
CC antiproliferative drug, inhibitor of lymphocyte activation, steroid,  
CC purine antimetabolite, antibody or immunomodulator. Accordingly, these  
CC immunosuppressive compositions can be used for treating or preventing  
CC graft rejection associated with transplantation of cells, tissue or  
CC organs (HLA-matched or mismatched) selected from hematopoietic cells,  
CC stem cells, heart, lung, kidney, liver or skin. The copolymer 1 or its  
CC related heteropolymer in combination with other immunosuppressive drugs  
CC increases the effectiveness of the drugs at lower dosages, reducing toxic  
CC side effects such that they are useful for host-versus-graft (HVG)  
CC response, as well as graft-versus-host disease (GVHD). This peptide  
CC sequence is an ordered copolymer of the invention.  
XX Sequence 15 AA;  
Query Match 87.0%; Score 47; DB 9; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 AFAAAAAAAAAA 13  
|:|||||  
DB 3 AYAAAAAAAAAA 14  
|:|||||

QY 2 AFAAAAAAAAAA 13  
|:|||||  
Db 3 AYAAAAAAAAAA 14

Search completed: September 9, 2006, 22:48:09  
Job time : 99.4177 secs

RESULT 40  
AEA35261  
ID AEA35261 standard; peptide; 15 AA.  
XX  
AC AEA35261;  
XX  
DT 28-JUL-2005 (first entry)  
XX  
DE Cop 1 binding motif #20.  
XX  
KW Anticonvulsant; Nootropic; Neuroprotective; Antiparkinsonian; Cop 1;  
KW Copolymer 1; HLA-DR; neurodegenerative disease; glutamate toxicity;  
KW Huntingtons disease; Alzheimers disease; Parkinsons disease.  
XX  
OS Synthetic.  
XX  
PN WO2005046719-A1.  
XX  
PD 26-MAY-2005.  
XX  
PF 11-NOV-2004; 2004WO-IL001037.  
XX  
PR 12-NOV-2003; 2003US-0518627P.  
PR 20-SEP-2004; 2004US-0610966P.  
XX  
PA (YEDA ) YEDA RES & DEV CO LTD.  
XX  
PI Eisenbach-Schwartz M, Yoles E, Butovsky O, Kipnis J;  
XX  
DR WPI; 2005-372283/38.  
XX  
PT Treating a neurodegenerative disorder or disease where there is  
PT accumulation of misfolded and/or aggregated proteins e.g. Parkinson's  
PT disease comprises using an agent e.g. Copolymer 1 or a Copolymer 1-  
PT related peptide.  
XX  
PS Disclosure; SEQ ID NO 20; 63pp; English.  
XX  
CC This peptide represents a peptide derived from Cop 1 (Copolymer 1). This  
CC peptide was tested for binding to the peptide binding groove of HLA-DR.  
CC The method of the invention for treating a neurodegenerative disorder or  
CC disease in which there is accumulation of misfolded and/or aggregated  
CC proteins, excluding prion-related diseases, comprises administering to an  
CC individual an agent selected from Copolymer 1, a Copolymer 1-related  
CC peptide, a Copolymer 1-related polypeptide, and T cells activated with  
CC them. The invention also includes methods for: reducing disease  
CC progression, and/or for protection from neurodegeneration and/or  
CC protection from glutamate toxicity in a patient suffering from a  
CC neurodegenerative disease or disorder selected from Huntington's disease,  
CC Alzheimer's disease and Parkinson's disease; and treating or preventing  
CC neurodegeneration and cognitive decline and dysfunction associated with  
CC Huntington's disease, Alzheimer's disease or Parkinson's disease.  
CC Specifically, treating a patient suffering from a neurodegenerative  
CC disease or disorder comprises immunizing the patient with a vaccine  
CC comprising Copolymer 1 for reducing disease progression or for protection  
CC from neurodegeneration in the patient. A pharmaceutical composition as  
CC administered in the method above, is useful for reducing disease  
CC progression, and/or for protection from neurodegeneration, and/or  
CC protection from glutamate toxicity in a patient suffering from the  
CC neurodegenerative disorder or disease, such as Huntington's disease,  
CC Alzheimer's disease or Parkinson's disease. The active agent is useful  
CC for manufacturing a medicament for treatment of neurodegenerative  
CC disorder or disease in which there is accumulation of misfolded and/or  
CC aggregated proteins, excluding prion-related diseases.

XX Sequence 15 AA;  
SQ  
Query Match 87.0%; Score 47; DB 9; Length 15;  
Best Local Similarity 91.7%; Pred. No. 1.6;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;



GenCore version 5.1.9  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.  
 OM protein - protein search, using sw model  
 Run on: September 9, 2006, 22:48:41 ; Search time 13.6709 Seconds  
 (without alignments)  
 84.457 Million cell updates/sec

Title: US-10-617-568-2  
 Perfect score: 50  
 Sequence: 1 AAAAAAAAAA 12

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR\_80.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	82.0	85	2	F84306
2	41	82.0	443	1	I38239
3	40	80.0	94	2	T03285
4	40	80.0	376	2	A26066
5	40	80.0	377	2	T04213
6	40	80.0	378	2	A44443
7	40	80.0	392	2	B48423
8	40	80.0	401	2	A48423
9	40	80.0	459	2	B44498
10	40	80.0	494	2	A42170
11	40	80.0	497	2	JC5076
12	40	80.0	509	2	A36392
13	40	80.0	530	2	I38558
14	40	80.0	577	2	T92227
15	40	80.0	892	2	T09071
16	40	80.0	1077	2	A44067
17	40	80.0	1533	2	A46221
18	40	80.0	2038	2	A43742
19	39	78.0	301	2	T35129
20	39	78.0	451	2	D88395
21	39	78.0	668	2	B96740
22	39	78.0	1596	2	A33106
23	39	78.0	1655	2	T13998
24	39	78.0	1668	2	T13748
25	38	76.0	40	2	S58853
26	38	76.0	85	1	FDFL4W
27	38	76.0	91	2	A22592
28	38	76.0	150	2	T12547
29	38	76.0	172	2	S35568

30	38	76.0	220	2	JC5954
31	38	76.0	231	2	S28186
32	38	76.0	233	2	S11563
33	38	76.0	238	2	A48279
34	38	76.0	289	2	A43562
35	38	76.0	305	2	I57039
36	38	76.0	314	2	JC5273
37	38	76.0	323	2	S16318
38	38	76.0	328	2	AB1781
39	38	76.0	331	2	A47236
40	38	76.0	333	2	A39065
41	38	76.0	334	2	G02409
42	38	76.0	364	2	I48188
43	38	76.0	374	2	T03875
44	38	76.0	375	2	T03874
45	38	76.0	403	2	A53662
46	38	76.0	420	2	T14911
47	38	76.0	425	2	A38153
48	38	76.0	448	2	S17370
49	38	76.0	475	2	A43915
50	38	76.0	477	2	A47236
51	38	76.0	482	2	JC7583
52	38	76.0	490	2	T09084
53	38	76.0	495	1	S31223
54	38	76.0	513	2	A48233
55	38	76.0	537	2	A55929
56	38	76.0	552	1	WJFFEN
57	38	76.0	604	2	A39369
58	38	76.0	606	2	S13367
59	38	76.0	627	2	T02610
60	38	76.0	630	2	T38023
61	38	76.0	640	2	A41726
62	38	76.0	642	2	S27806
63	38	76.0	644	2	S39356
64	38	76.0	649	2	S43229
65	38	76.0	701	1	S46458
66	38	76.0	702	2	G01840
67	38	76.0	779	1	S40382
68	38	76.0	792	1	EAHU
69	38	76.0	796	2	JC7555
70	38	76.0	799	2	JH0797
71	38	76.0	805	2	T49385
72	38	76.0	806	2	T13690
73	38	76.0	828	2	C88402
74	38	76.0	829	2	A34692
75	38	76.0	830	2	A80369
76	38	76.0	873	2	B53225
77	38	76.0	883	2	S04722
78	38	76.0	1028	2	A56038
79	38	76.0	1052	2	T14343
80	38	76.0	1065	2	T13230
81	38	76.0	1072	2	T13232
82	38	76.0	1074	2	T13229
83	38	76.0	1081	2	T13231
84	38	76.0	1137	2	A33507
85	38	76.0	1180	2	S69205
86	38	76.0	1205	2	A55015
87	38	76.0	1212	2	A57187
88	38	76.0	1213	2	S16356
89	38	76.0	1355	2	S40022
90	38	76.0	1421	2	T49500
91	38	76.0	1430	2	T34516
92	38	76.0	1506	2	T31900
93	38	76.0	1647	2	T49412
94	38	76.0	1891	2	T13594
95	38	76.0	1920	2	T13893
96	38	76.0	2175	1	S03170
97	38	76.0	2639	2	T31328
98	38	76.0	2715	2	T13049
99	38	76.0	2957	2	T33152
100	38	76.0	5327	2	T13564

ribosomal protein  
 achete-scute locu  
 probable MASH-2 pr  
 achete scute prot  
 homeotic protein H  
 genomic screen hom  
 paired type homeob  
 homeotic protein H  
 hypothetical protei  
 zinc-finger protei  
 homeotic protein E  
 protein kinase C-b  
 Gene NKx6.1 protei  
 probable homeobox  
 homeotic protein H  
 bZIP DNA-binding p  
 paired box protein  
 DNA-binding protei  
 homeotic protein e  
 zinc-finger protei  
 basic helix-loop-h  
 phosphatidylinosit  
 transcription fact  
 polyomavirus enhan  
 zinc finger protei  
 homeotic protein e  
 homeotic protein B  
 Om(1D) protein - f  
 probable YME1 ATP-  
 probable transcrip  
 homeotic protein B  
 homeotic protein B  
 homeotic protein B  
 transcription fact  
 arylsulfatase (EC  
 transcription fact  
 T-box protein 2 -  
 box A-binding fact  
 elastin precursor,  
 Cl4orf4 protein -  
 castor protein - f  
 hypothetical prote  
 hypothetical protei  
 protein H05C05.1 f  
 ecdysone-induced p  
 nitrate reductase  
 ecdysone-induced p  
 puff 74E protein -  
 DNA-binding protei  
 zinc finger RNA bi  
 dachshund isoform  
 dachshund protein  
 dachshund protein  
 dachshund protein  
 stripe a/b protein  
 hypothetical prote  
 bumetanide-sensiti  
 bumetanide-sensiti  
 ovo protein - frui  
 spalt protein - fr  
 hypothetical prote  
 hypothetical prote  
 related to PAN2 pr  
 hypothetical prote  
 hypothetical prote  
 gene hindsight pro  
 homeotic protein c  
 fibrocin - Chinese  
 eyelid - fruit fly  
 hypothetical prote  
 microtubule-associ

## ALIGNMENTS

RESULT 1  
F84306  
hypothetical protein Vng1530h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: F84306  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabcic, Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: F84306  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-85 <STO>  
A:Cross-references: UNIPROT:Q9HPP9; UNIPARC:UPI00000638F2; GB:AE004437; NID:g10581018; F  
C:Genetics:  
A:Gene: VNG1530H

Query Match 82.0%; Score 41; DB 2; Length 85;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 10  
|||  
Db 30 AAAAAAAAAA 39

RESULT 2  
I38239  
transcription factor SOX3 - human  
N:Alternate names: SRY (sex determining region Y)-box 3  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: I38239; I38242; S67816  
R:Stevanovic, M.; Lovell-Badge, R.; Collignon, J.; Goodfellow, P.N.  
Hum. Mol. Genet. 2, 2013-2018, 1993  
A:Title: SOX3 is an X-linked gene related to SRY.  
A:Reference number: I38239; MUID:94154672; PMID:8111369  
A:Accession: I38239  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-443 <STE1>  
A:Cross-references: UNIPROT:P41225; UNIPARC:UPI000003F546; EMBL:X71135; NID:g468790; PID  
A:Accession: I38242  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 143-158, 'P', 160-218 <STE2>  
A:Cross-references: UNIPARC:UPI0000071C8C; EMBL:X71137; NID:g468793; PIDN:CAA50467.1; PI  
C:Genetics:  
A:Gene: GDB:SOX3; SOX-3; SOXB  
A:Cross-references: GDB:250376; OMIM:313430  
A:Map position: Xq26-Xq27  
C:Superfamily: human SOX3 protein; HMG box homology  
F:136-211/Domain: HMG box homology <HMG>

Query Match 82.0%; Score 41; DB 1; Length 443;  
Best Local Similarity 83.3%; Pred. No. 66;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
|||  
Db 352 AAAAAAAAAA 363

RESULT 3  
T03285  
anther-specific protein - rice

C:Species: Oryza sativa (rice)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C:Accession: T03285  
R:Lee, J.Y.K.; Hodges, T.K.  
submitted to the EMBL Data Library, July 1994  
A:Description: Genomic DNA sequence of a rice anther-specific gene.  
A:Reference number: Z14882  
A:Accession: T03285  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-94 <LEE>  
A:Cross-references: UNIPROT:Q40629; UNIPARC:UPI00000A8A96; EMBL:U12171; NID:g607894; PID  
A:Experimental source: strain IR54  
C:Genetics:  
A:Gene: RTS2

Query Match 80.0%; Score 40; DB 2; Length 94;  
Best Local Similarity 90.9%; Pred. No. 27;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAA 12  
|||  
Db 14 AAAAAAAAAA 24

RESULT 4  
A26066  
segmentation protein eve - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 25-Oct-1987 #sequence\_revision 25-Oct-1987 #text\_change 31-Dec-2004  
C:Accession: A26066; B26066; A26636  
R:Macdonald, P.M.; Ingham, P.; Struhl, G.  
Cell 47, 721-734, 1986  
A:Title: Isolation, structure, and expression of even-skipped: a second pair-rule gene  
A:Reference number: A26066; MUID:87051744; PMID:2877745  
A:Accession: A26066  
A:Molecule type: DNA  
A:Residues: 1-376 <MAC>  
A:Cross-references: UNIPROT:P06602; UNIPARC:UPI000016BBB4; GB:M14767; NID:g157386; PIDN:  
A:Accession: B26066  
A:Molecule type: mRNA  
A:Residues: 32-268, 279-376 <MA2>  
A:Cross-references: UNIPARC:UPI000017A27C; UNIPARC:UPI000017A27D; GB:M14767  
R:Frasch, M.; Hoey, T.; Rushlow, C.; Doyle, H.; Levine, M.  
EMBO J. 6, 749-759, 1987  
A:Title: Characterization and localization of the even-skipped protein of Drosophila.  
A:Reference number: A26636; MUID:87218536; PMID:2884106  
A:Accession: A26636  
A:Molecule type: mRNA  
A:Residues: 1-299, 'L', 301-376 <FRA>  
A:Cross-references: UNIPARC:UPI0000124746; GB:X05138; NID:g7957; PIDN:CAA28784.1; PID:g  
C:Genetics:  
A:Gene: eve (even-skipped)  
A:Cross-references: FlyBase:FBgn0000606  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:71-127/Domain: homeobox homology <HOX>

Query Match 80.0%; Score 40; DB 2; Length 376;  
Best Local Similarity 83.3%; Pred. No. 79;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
|||  
Db 168 AAAAAAAAAA 179

RESULT 5  
T04213  
heat shock transcription factor homolog T5C23.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T04213  
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15261  
A:Accession: T04213  
A:Molecule type: DNA  
A:Residues: 1-377 <BEV>  
A:Cross-references: UNIPROT:Q9TOD3; UNIPARC:UPI000000117B; EMBL:AL049500  
A:Experimental source: cultivar Columbia; BAC clone T5C23  
C:Genetics:  
A:Map position: 4  
A:Introns: 118/3  
A:Note: T5C23.90  
F:60-155/Domain: HSF DNA-binding domain homology <HSF>

Query Match 80.0%; Score 40; DB 2; Length 377;  
Best Local Similarity 90.9%; Pred. No. 79;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAAAAAAAAA 12  
||| |||||  
Db 155 AAAAAAAAAA 165

## RESULT 6

A4443 basic helix-loop-helix transcription repressor H - fruit fly (Drosophila virilis)

C:Species: Drosophila virilis  
C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Jul-1998  
C:Accession: A4443  
R:Wainwright, S.M.; Ish-Horowicz, D.  
Mol. Cell. Biol. 12, 2475-2483, 1992  
A:Title: Point mutations in the Drosophila hairy gene demonstrate in vivo requirements for enhancer activity  
A:Reference number: A4443; MUID:92269819; PMID:1588951  
A:Accession: A4443  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-378 <WAI>  
A:Cross-references: UNIPARC:UPI000017BEF1  
A:Note: sequence inconsistent with the nucleotide translation  
C:Genetics:  
A:Gene: FlyBase:Dvir/h  
A:Cross-references: FlyBase:FBgn0013115

Query Match 80.0%; Score 40; DB 2; Length 378;  
Best Local Similarity 83.3%; Pred. No. 79;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
||| |||||  
Db 272 AAAAAAAAAA 283

## RESULT 7

B48423 homeotic protein engrailed 1 - human  
C:Species: Homo sapiens (man)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 05-Oct-2004  
C:Accession: B48423  
R:Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.L.  
Dev. Genet. 13, 345-358, 1992  
A:Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene  
A:Reference number: A48423; MUID:93185339; PMID:1363401  
A:Accession: B48423  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-392 <LOG>  
A:Cross-references: UNIPROT:Q05925; UNIPARC:UPI000012CA02  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:304-360/Domain: homeobox homology <HOX>

Query Match 80.0%; Score 40; DB 2; Length 392;  
Best Local Similarity 83.3%; Pred. No. 82;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
||| |||||  
Db 199 AAAAAAAAAA 210

## RESULT 8

A48423 engrailed homeodomain-containing protein En-1 - mouse  
N:Alternate names: homeotic protein En-1  
C:Species: Mus musculus (house mouse)  
C>Date: 01-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C:Accession: A48423; SI3009; A26629; A24778  
R:Logan, C.; Hanks, M.C.; Noble-Topham, S.; Nallainathan, D.; Provart, N.J.; Joyner, A.L.  
Dev. Genet. 13, 345-358, 1992  
A:Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene  
A:Reference number: A48423; MUID:93185339; PMID:1363401  
A:Accession: A48423  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-401 <LOG>  
A:Cross-references: UNIPROT:P09065; UNIPARC:UPI00000299A1  
A:Experimental source: CD-1, embryo  
A:Note: sequence extracted from NCBI backbone (NCBI:126620)  
R:Holland, P.W.H.; Williams, N.A.  
FEBS Lett. 277, 250-252, 1990  
A:Title: Conservation of engrailed-like homeobox sequences during vertebrate evolution.  
A:Reference number: SI3009; MUID:91099509; PMID:1980115  
A:Accession: SI3009  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 321-380 <HOL>  
A:Cross-references: UNIPARC:UPI000017A2AF  
R:Joyner, A.L.; Martin, G.R.  
Genes Dev. 1, 29-38, 1987  
A:Title: En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrailed gene  
A:Reference number: A91620; MUID:88112776; PMID:2892757

A:Accession: A26629

A:Molecule type: DNA; mRNA

A:Residues: 278-401 <JOY>

A:Cross-references: UNIPARC:UPI000016CA7C; GB:Y00201; GB:M11987; NID:G49587; PIDN:CAA68

R:Joyner, A.L.; Kornberg, T.; Coleman, K.G.; Cox, D.R.; Martin, G.R.  
Cell 43, 29-37, 1985

A:Title: Expression during embryogenesis of a mouse gene with sequence homology to the

A:Reference number: A24778; MUID:86079501; PMID:2416459

A:Accession: A24778

A:Molecule type: DNA

A:Residues: 311-401 <JO2>

A:Cross-references: UNIPARC:UPI000017A2B0

C:Genetics:

A:Gene: en.1

A:Map position: 1

C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

F:313-369/Domain: homeobox homology <HOX>

Query Match 80.0%; Score 40; DB 2; Length 401;

Best Local Similarity 83.3%; Pred. No. 83;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
||| |||||  
Db 211 AAAAAAAAAA 222

## RESULT 9

B44498 radial spoke protein 6 - Chlamydomonas reinhardtii

C:Species: Chlamydomonas reinhardtii  
C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Oct-2004  
C:Accession: B44498  
R:Curry, A.M.; Williams, B.D.; Rosenbaum, J.L.  
Mol. Cell. Biol. 12, 3967-3977, 1992  
A:Title: Sequence analysis reveals homology between two proteins of the flagellar radial

A:Reference number: A44498; MUID:92375065; PMID:1508197

A:Accession: B44498

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-459 <CR>

A:Cross-references: UNIPROT:Q01657; UNIPARC:UPI000013521B; GB:M87526; NID:g167435; PIDN:

A:Note: sequence extracted from NCBI backbone (NCBIN:111696, NCBIP:111698)

C:Superfamily: flagellar radial spoke protein

Query Match 80.0%; Score 40; DB 2; Length 459;

Best Local Similarity 83.3%; Pred. No. 92;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12

|| |||||:|

Db 66 AASAAAAA 77

RESULT 10

A42170

zinc finger protein MAZ - human (fragment)

N:Alternate names: MYC-associated zinc finger protein MAZ; zinc finger protein ZF87

C:Species: Homo sapiens (man)

C:Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 09-Jul-2004

C:Accession: A42170; A46153

R:Pyrc, J.J.; Moberg, K.H.; Hall, D.J.

Biochemistry 31, 4103-4110, 1992

A:Title: Isolation of a novel cDNA encoding a zinc-finger protein that binds to two site

A:Reference number: A42170; MUID:92232709; PMID:1567856

A:Accession: A42170

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-494 <PR>

A:Cross-references: UNIPROT:P56270; UNIPARC:UPI000017C427; GB:J05371

A:Note: it is uncertain whether Met-18 is the initiator or whether translation is initia

R:Bossonne, S.A.; Asselin, C.; Patel, A.J.; Marcu, K.B.

Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456, 1992

A:Title: MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating tra

A:Reference number: A46153; MUID:92366479; PMID:1502157

A:Accession: A46153

A:Molecule type: mRNA

A:Residues: 18-417, 'L', 419-494 <BOS>

A:Cross-references: UNIPARC:UPI000012ECF8; GB:M94046

A:Experimental source: HeLa cells

A:Note: sequence extracted from NCBI backbone (NCBIN:110666, NCBIP:110667)

C:Keywords: DNA binding; zinc finger

F:113-125/Region: alanine-rich

F:174-183/Region: alanine-rich

F:207-230/Region: zinc finger

F:296-318/Region: zinc finger

F:324-346/Region: zinc finger

F:354-368/Region: zinc finger

F:373-405/Region: zinc finger

F:409-430/Region: zinc finger

F:452-468/Region: alanine-rich

Query Match 80.0%; Score 40; DB 2; Length 494;

Best Local Similarity 83.3%; Pred. No. 98;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12

|| |||||:|

Db 116 AAAAAAAAAA 127

RESULT 11

JC5076

MYC-associated zinc-finger protein - human

N:Alternate names: MAZ protein

C:Species: Homo sapiens (man)

C:Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 05-Nov-1999

C:Accession: JC5076

R:Tsutsui, H.; Sakatsume, O.; Itakura, K.; Yokoyama, K.K.

Biochem. Biophys. Res. Commun. 226, 801-809, 1996

A:Title: Members of the MAZ family: A novel cDNA clone for MAZ from human pancreatic isl

A:Reference number: JC5076; MUID:96428591; PMID:8831693

A:Accession: JC5076

A:Molecule type: mRNA

A:Residues: 1-497 <TSU>

A:Cross-references: UNIPARC:UPI0000163B39; DDBJ:D85131; NID:g1752741; PIDN:BA412728.1; P

A:Experimental source: pancreatic islet

C:Comment: This protein plays a role in the control of transcriptional initiation of gen

and between the introns of the mouse gene for immunoglobulin M-D.

C:Keywords: phosphoprotein; zinc finger

F:146,204,480/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status pre

F:349/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 80.0%; Score 40; DB 2; Length 497;

Best Local Similarity 83.3%; Pred. No. 98;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12

|| |||||:|

Db 125 AAAAAAAAAA 136

RESULT 12

A36392

segmentation protein runt - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 09-Jul-2004

C:Accession: A36392

R:Kania, M.A.; Bonner, A.S.; Duffy, J.B.; Gergen, J.P.

Genes Dev. 4, 1701-1713, 1990

A:Title: The Drosophila segmentation gene runt encodes a novel nuclear regulatory protei

A:Reference number: A36392; MUID:91065517; PMID:2249771

A:Accession: A36392

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-509 <KAN>

A:Cross-references: UNIPROT:P22814; UNIPARC:UPI0000135332; GB:X56432; NID:g8521; PIDN:CA

C:Genetics:

A:Gene: FlyBase:run

A:Cross-references: FlyBase:FBgn0003300

A:Superfamily: transcription factor CBF alpha 2

C:Keywords: transcription regulation

Query Match 80.0%; Score 40; DB 2; Length 509;

Best Local Similarity 83.3%; Pred. No. 1e+02;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12

|| |||||:|

Db 20 AAAAAAAAAA 31

RESULT 13

I38558

Mi-2 autoantigen 240 kDa protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 05-Nov-1999

C:Accession: I38558

R:Ge, Q.; Nilasena, D.S.; O'Brien, C.A.; Frank, M.B.; Targoff, I.N.

J. Clin. Invest. 96, 1730-1737, 1995

A:Title: Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 au

A:Reference number: I38558; MUID:96013633; PMID:7560064

A:Accession: I38558

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-530 <RES>

A:Cross-references: UNIPARC:UPI000016A09A; EMBL:U08379; NID:g761717; PIDN:AAC50228.1; P

Query Match

Best Local Similarity 83.3%; Score 40; DB 2; Length 530;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
 ||:|||||||  
 Db 87 AAAAAAAAAA 98

## RESULT 14

S72227  
 finger protein sob - fruit fly (Drosophila melanogaster)  
 C/Species: Drosophila melanogaster  
 C/Date: 04-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 09-Jul-2004  
 C/Accession: S72227  
 R/Hart, M.C.; Wang, L.; Coulter, D.E.  
 Genetics 144, 171-182, 1996  
 A/Title: Comparison of the structure and expression of odd-skipped and two related genes  
 A/Reference number: S72227; MUID:97032935; PMID:8878683  
 A/Accession: S72227  
 A/Molecule type: mRNA  
 A/Status: preliminary  
 A/Residues: 1-577 <HAR>  
 A/Cross-references: UNIPROT:Q24571; UNIPARC:UPI000007A28D; EMBL:U62004; NID:g1480193; PI  
 A/Gene: sob  
 C/Genetics:  
 A/Cross-references: FlyBase:FBgn0004892  
 C/Keywords: zinc finger

Query Match 80.0%; Score 40; DB 2; Length 577;  
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
 ||:|||||||  
 Db 273 AAAAAAAAAA 284

## RESULT 15

T09071  
 SH3 domains-containing protein POSH - mouse  
 C/Species: Mus musculus (house mouse)  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
 C/Accession: T09071  
 R/Tapon, N.; Nagata, K.; Lamarche, N.; Hall, A.  
 EMBO J. 17, 1395-1404, 1998  
 A/Reference number: Z16552; MUID:98151363; PMID:9482736  
 A/Accession: T09071  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-892 <TAP>  
 A/Cross-references: UNIPROT:O70254; UNIPARC:UPI0000027BF6; EMBL:AF030131; NID:g3002587;  
 A/Experimental source: cell line Ras-transformed NIH 3T3 cells  
 A/Note: activates JNK/SAPK cascade; Rac-binding protein  
 C/Genetics:  
 A/Gene: POSH  
 C/Keywords: signal transduction  
 F;8-58/Domain: RING finger homology <RRN>

Query Match 80.0%; Score 40; DB 2; Length 892;  
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
 ||:|||||||  
 Db 419 AAAAAAAAAA 430

## RESULT 16

A44067  
 serine-rich protein hairless - fruit fly (Drosophila melanogaster)  
 N/Alternate names: 109K basic protein H  
 C/Species: Drosophila melanogaster  
 C/Date: 10-Jun-1993 #sequence\_revision 26-Feb-1999 #text\_change 09-Jul-2004  
 C/Accession: A44067; A58929; S33412; S24639  
 R/Bang, A.G.; Posakony, J.W.  
 Genes Dev. 6, 1752-1769, 1992

A/Title: The Drosophila gene Hairless encodes a novel basic protein that controls alter  
 A/Reference number: A44067; MUID:92387549; PMID:1516831  
 A/Accession: A44067  
 A/Molecule type: DNA  
 A/Residues: 19-1077 <BAN>  
 A/Cross-references: UNIPROT:Q02308; UNIPARC:UPI000002AB24; GB:M95192; NID:g157621; PID:  
 A/Note: sequence extracted from NCBI backbone (NCBIN:112622, NCBIPI:112623)  
 R/Preiss, A.

submitted to the EMBL Data Library, May 1994  
 A/Description: Hairless, a Drosophila gene involved in neural development, encodes a novel  
 A/Reference number: A58929

A/Accession: A58929  
 A/Molecule type: mRNA  
 A/Residues: 1-1077 <PRE>  
 A/Cross-references: UNIPARC:UPI000012C93C; EMBL:XG7239; GB:S49642; NID:g578331; PID:g57  
 R/Mater, D.; Stumm, G.; Kuhn, K.; Preiss, A.  
 Mech. Dev. 38, 143-156, 1992

A/Title: Hairless, a Drosophila gene involved in neural development, encodes a novel,  
 A/Reference number: S33412; MUID:93041287; PMID:1419850

A/Accession: S33412  
 A/Molecule type: mRNA  
 A/Residues: 1-150,'A',152-701,'LL',704-890,'R',892-963,'RLLP',968-973,975-1077 <NAI>  
 A/Cross-references: UNIPARC:UPI000017BED1; EMBL:XG7239  
 C/Genetics:

A/Gene: FlyBase:H; hairless  
 A/Cross-references: FlyBase:FBgn00001169

Query Match 80.0%; Score 40; DB 2; Length 1077;  
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
 ||:|||||||  
 Db 880 AAAAAAAAAA 891

## RESULT 17

A46221  
 abdominal segment formation protein pumilio - fruit fly (Drosophila melanogaster)  
 C/Species: Drosophila melanogaster  
 C/Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C/Accession: A46221; S22026  
 R/Barker, D.D.; Wang, C.; Moore, J.; Dickinson, L.K.; Lehmann, R.  
 Genes Dev. 6, 2312-2326, 1992  
 A/Title: Pumilio is essential for function but not for distribution of the Drosophila al  
 A/Reference number: A46221; MUID:93093466; PMID:1459455  
 A/Contents: embryo  
 A/Accession: A46221  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-1533 <BAR>  
 A/Cross-references: UNIPROT:P25822; UNIPARC:UPI000016BD14; GB:L07943; NID:g158190; PID:  
 A/Note: sequence extracted from NCBI backbone (NCBIN:120203, NCBIPI:120204)  
 R/Macdonald, P.M.  
 submitted to the EMBL Data Library, October 1991

A/Reference number: S22026  
 A/Accession: S22026  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-361,'A',363-1102,'R',1104-1405,'KN',1408-1495,'V',1497-1518,'S',1520-1533  
 A/Cross-references: UNIPARC:UPI00001329FD; EMBL:X62589; NID:g8393; PID:g8394  
 C/Genetics:  
 A/Gene: FlyBase:pum  
 A/Cross-references: FlyBase:FBgn0003165

Query Match 80.0%; Score 40; DB 2; Length 1533;  
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
 ||:|||||||  
 Db 1050 AAAAAAAAAA 1061

```
RESULT 18
A43742
female sterile homeotic protein, 205K - fruit fly (Drosophila melanogaster)
N:Alternate names: membrane protein fish, 205K
N:Contains: female sterile homeotic protein, 110K
C:Species: Drosophila melanogaster
C>Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 31-Dec-2004
C:Accession: A43742; B43742
R:Haynes, S.R.; Mozer, B.A.; Bhatia-Bey, N.; Dawid, I.B.
Dev. Biol. 134, 246-257, 1989
A:Title: The Drosophila fish locus, a maternal effect homeotic gene, encodes apparent men
A:Reference number: A43742; MUID:89276730; PMID:2567251
A:Accession: A43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2038 <HAY>
A:Cross-references: UNIPROT:P13709; UNIPARC:UPI000012AC6C; EMBL:M23221; NID:G157452; PID
A:Accession: B43742
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1106 <HA2>
A:Cross-references: UNIPARC:UPI000017A10F; EMBL:M23222
C:Genetics:
A:Gene: fish
A:Cross-references: FlyBase:FBgn0004656
C:Keywords: alternative splicing; transmembrane protein
F:1-2038/Product: female sterile homeotic protein, 205K #status predicted <MA2>
F:1-1106/Product: female sterile homeotic protein, 110K #status predicted <MA1>
F:59-116/Domain: bromodomain homology <BRO1>
F:503-560/Domain: bromodomain homology <BRO2>

Query Match 80.0%; Score 40; DB 2; Length 2038;
Best Local Similarity 83.3%; Pred. No. 3e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAMAAAAAAAMA 12
||| |||||
Db 330 AAVAAAAAAAMA 341

RESULT 19
T35129
probable hydrolase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004
C:Accession: T35129
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z21568
A:Accession: T35129
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-301 <SEE>
A:Cross-references: UNIPROT:Q69985; UNIPARC:UPI00000DAC0B; EMBL:AL022268; PIDN:CAA18345.
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SC0EDB:SC4H2.30
C:Superfamily: ADP-ribosylglycohydrolase

Query Match 78.0%; Score 39; DB 2; Length 301;
Best Local Similarity 81.8%; Pred. No. 90;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AMAAAAAAAMA 12
||| |||||
Db 165 AMAVAAAAALA 175

RESULT 20
D88395
protein F53A3.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
```

```
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: D88395
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D88395
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-451 <STO>
A:Cross-references: UNIPROT:O17222; UNIPARC:UPI000017A5BC; GB:chr_III; PIDN:AAB70991.1;
C:Genetics:
A:Gene: F53A3.6
A:Map position: 3

Query Match 78.0%; Score 39; DB 2; Length 451;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAMAAAAAAAMA 12
||| |||||
Db 303 AAQAAAAAAAMA 314

RESULT 21
B96740
hypothetical protein F14023.15 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: B96740
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B96740
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-668 <STO>
A:Cross-references: UNIPROT:Q05196; UNIPARC:UPI00001311AE; GB:AE005173; NID:G7239504; PT
C:Genetics:
A:Gene: F14023.15
A:Map position: 1
C:Superfamily: polyadenylate-binding protein; ribonucleoprotein repeat, homology

Query Match 78.0%; Score 39; DB 2; Length 668;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAMAAAAAAAMA 12
||| |||||
Db 27 AAVAAAAAAAEA 38

RESULT 22
A33106
neurogenic locus mam protein - fruit fly (Drosophila melanogaster)
N:Alternate names: mastermind protein
C:Species: Drosophila melanogaster
C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: A36391; A33106; S13514
R:Smoller, D.; Friedel, C.; Schmid, A.; Bettler, D.; Lam, L.; Yedvobnick, B.
Genes Dev. 4, 1688-1700, 1990
A:Title: The Drosophila neurogenic locus mastermind encodes a nuclear protein unusually
A:Reference number: A36391; MUID:91065516; PMID:1701150
```

A:Accession: A36391  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1596 <SMO>  
A:Cross-references: UNIPROT  
A:Note: strain Canton S  
C:Genetics:  
A:Gene: FlyBase:mam  
A:Cross-references: FlyBase

Query Match	78.0%;	Score 39;	DB 2;	Length 1596;
Best Local Similarity	83.3%;	Pred. No.	3.3e+02;	
Matches 10;	Conservative	0;	Mismatches 2;	Indels 0;
Gaps	0;			

QY 1 AAAAAAAAAA 12  
|||  
Db 1062 AAAAAAAAAA 1073

RESULT 23  
Tl3998  
gene mastermind protein - fruit fly (*Drosophila virilis*)  
C:Species: *Drosophila virilis*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: Tl3998  
R:Newfield, S.J.; Tachida, H.; Yedvobnick, B.  
J. Mol. Evol. 38, 637-641, 1994  
A:Title: Drive-selection equilibrium: homopolymer evolution in the *Drosophila* gene mastermind  
A:Reference number: Z17850; MUID:94365848; PMID:8083889  
A:Accession: Tl3998  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-1655 <NEW>  
A:Cross-references: UNIPROT:Q24754; UNIPARC:UPI00000078F11; EMBL:M92914; NID:g157833; PIR:G157833  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0013119

Query Match	78.0%	Score 39;	DB 2;	Length 1655;
Best Local Similarity	83.3%	Pred. No.	3.4e+02;	
Matches 10;	Conservative	0;	Mismatches 2;	Indels 0;
Gaps 0;				

QY 1 AAAAAAAAAA 12  
Db 1103 AAAAAAAAAA 1114

RESULT 24  
Tl3748  
sex comb protein - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: Tl3748  
E:Sinclair, D.A.R.; Milne, T.A.; Hodgson, J.W.; Shellard, J.; Salinas, C.A.; Kyba, M.; R  
Development 125, 1207-1216, 1998  
A:Title: The Additional sex combs gene of *Drosophila* encodes a chromatin protein that bi  
A:Reference number: Tl7750; MUID:98146384; PMID:9477319  
A:Accession: Tl3748  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1668 <SIN>  
A:Cross-references: UNIPARC:UPI00000760FA; EMBL:AJ001164; NID:g3292938; PIDN:CAA04568.1;  
C:Genetics:  
A:Cross-references: FlyBase:FBgn0000142  
C:Function:  
A:Description: involved in repression of homeotic loci

Query Match 78.0%; Score 39; DB 2; Length 1668;  
Best Local Similarity 83.3%; Pred. No. 3.5e+02;  
Matches 10: Conservative 0; Mismatches 2; Indels

Qy 1 AAAAAAAAAA 12  
||| ||||| |  
Db 139 AAAAAAAAAA 150

RESULT 25  
S58853  
homeotic protein ultrabithorax homolog - Junonia coenia (fragment)  
N:Alternate names: ultrabithorax homeodomain protein  
C:Species: Junonia coenia  
C:Date: 19-Mar-1997 #sequence\_revision 29-Aug-1997 #text\_change 05-Oct-2004  
C:Accession: S58853  
R:Warren, R.W.; Nagy, L.; Selegue, J.; Gates, J.; Carroll, S.  
Nature 372, 458-461, 1994  
A:Title: Evolution of homeotic gene regulation and function in flies and butterflies  
A:Reference number: S58850; MUID:95075456; PMID:7840822  
A:Accession: S58853  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-40 <WAR>  
A:Cross-references: UNIPROT:Q25210; UNIPARC:UPI000017A2FE; EMBL:L42137  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995  
C:Keywords: DNA binding; nucleosome; transcription regulation

Query Match 76.0%; Score 38; DB 2; Length 40;  
Best Local Similarity 83.3%; Pred. No. 25;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
23 AAAAAAAAAA 34

```

RESULT 26
FDL14W
antifreeze protein 4 precursor - winter flounder
C/Species: Pseudopleuronectes americanus (winter flounder)
C/Date: 01-Sep-1981 #sequence_revision 01-Sep-1981 #text_change 09-Jul-2004
C/Accession: A03193
R/Lin, Y.; Gross, J.K.
Proc. Natl. Acad. Sci. U.S.A. 78, 2825-2829, 1981
A>Title: Molecular cloning and characterization of winter flounder antifreeze
A/Reference numbers: A03193; MUID:81247379; PMID:6265915
A/Accession: A03193
A/Molecule type: mRNA
A/Residues: 1-85 <LIN>
A/Cross-references: UNIPROT:P02734; UNIPARC:UPI0000174278
C/Superfamily: antifreeze protein
C/Keywords: antifreeze
F/1-21/Domain: signal sequence #status predicted <Sig>
F/23-85/Product: antifreeze protein 4 #status predicted <MAT>

```

Query Match 76.0%; Score 38; DB 1; Length 85;  
Best Local Similarity 83.3%; Pred. No. 45;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 A A M A A A A A A A A A 12  
Dy 35 A A T A A A A A A A A T A 46

RESULT 27  
A22592  
antifreeze protein IIA7 precursor - winter flounder  
C:Species: Pseudopleuronectes americanus (winter flounder)  
C;Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 09-Jul-2004  
C;Accession: A22592; S12604  
R;Gourlie, B.; Lin, Y.; Price, J.; De Vries, A.L.; Powers, D.; Huang, R.C.C  
J. Biol. Chem. 259, 14960-14965, 1984  
A;Title: Winter flounder antifreeze proteins: a multigene family.  
A;Reference number: A22592; MUID:85054993; PMID:6548752  
A;Accession: A22592  
A;Molecule type: mRNA  
A;Residues: 1-91 <GOU>  
A;Cross-references: UNIPROT:P07835; UNIPARC:UPI0000125B48; GB:M10148; NID:G  
R;Gauthier, S.; Wu, Y.; Davies, P.L.

A;Residues: 1-91 <GOU>  
A;Cross-references: UNIPROT:P07835; UNIPARC:UPI0000125B4B; GB:M10148; NID:G213579; PIDN:  
R;Gauthier, S.; Wu, Y.; Davies, F.L.

Nucleic Acids Res. 18, 5303, 1990  
A:Title: Nucleotide sequence of a variant antifreeze protein gene.  
A:Reference number: S12604; MUID:90384854; PMID:2402466  
A:Accession: S12604  
A:Molecule type: DNA  
A:Residues: 1-63, 'V', 65-91 <GAU>  
A:Cross-references: UNIPARC:UPI0000001382; EMBL:X53718; NID:g64213; PIDN:CAA37754.1; PII  
C:Genetics:  
A:Introns: 19/2  
C:Superfamily: antifreeze protein  
C:Keywords: antifreeze; tandem repeat  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-91/Product: antifreeze protein IIA7 #status predicted <MAT>  
  
Query Match 76.0%; Score 38; DB 2; Length 91;  
Best Local Similarity 83.3%; Pred. No. 48;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAA 12  
Db 52 AATAAAAAAAAA 63  
  
RESULT 28  
T12547  
hypothetical protein DKFZp586E1621.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T12547  
R:Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z17528  
A:Accession: T12547  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-150 <GT>  
A:Cross-references: UNIPROT:Q9Y4M1; UNIPARC:UPI00000071DDD; EMBL:AL080235  
A:Experimental source: adult uterus; clone DKFZp586E1621  
C:Genetics:  
A:Note: DKFZp586E1621.1  
  
Query Match 76.0%; Score 38; DB 2; Length 150;  
Best Local Similarity 83.3%; Pred. No. 70;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAA 12  
Db 129 AAAAAAAAAA 140  
  
RESULT 29  
S35568  
sex-determining protein Sry - multimammate rat (Mastomys hildebrandtii)  
C:Species: Mastomys hildebrandtii  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Oct-2004  
C:Accession: S35568  
R:Tucker, P.K.; Lundrigan, B.L.  
Nature 364, 715-717, 1993  
A:Title: Rapid evolution of the sex determining locus in Old World mice and rats.  
A:Reference number: S35565; MUID:93361118; PMID:8355784  
A:Accession: S35568  
A:Molecule type: DNA  
A:Residues: 1-172 <TUC>  
A:Cross-references: UNIPROT:Q60586; UNIPARC:UPI000000671C; GB:L29542; NID:g496161; PIDN:  
C:Genetics:  
A:Gene: Sry  
A:Map position: Y  
C:Keywords: DNA binding  
F:2-77/Domain: HMG box homology <HMG1>  
  
Query Match 76.0%; Score 38; DB 2; Length 172;  
Best Local Similarity 83.3%; Pred. No. 79;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
Db 146 AAAAAAAAAA 157

#### RESULT 30

JC5954  
ribosomal protein L14 - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 02-Sep-2000  
C:Accession: JC5954  
R:Tanaka, M.; Tanaka, T.; Harata, M.; Suzuki, T.; Mitsui, Y.  
Biochem. Biophys. Res. Commun. 243, 531-537, 1998  
A:Title: Triplet repeat-containing ribosomal protein L14 gene in immortalized human endo  
A:Reference number: JC5954; MUID:96153799; PMID:9480843  
A:Accession: JC5954  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-220 <TAN>  
A:Cross-references: UNIPARC:UPI0000004E622; DBJ:D87735; NID:gi620021; PIDN:BAAL3443.1; P  
C:Superfamily: rat ribosomal protein L14

Query Match 76.0%; Score 38; DB 2; Length 220;  
Best Local Similarity 83.3%; Pred. No. 95;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
Db 150 AAAAAAAAAA 161

#### RESULT 31

S28186  
achaete-scute locus protein homolog MASH-1 - mouse  
N:Alternate names: mammalian achaete-scute homolog 1  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S28186  
R:del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Gridley, T.  
Biochim. Biophys. Acta 1171, 323-327, 1993  
A:Title: Cloning, sequencing and expression of the mouse mammalian achaete-scute homolog  
A:Reference number: S28186; MUID:93144349; PMID:8424959  
A:Accession: S28186  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-231 <AMO>  
A:Cross-references: UNIPROT:Q02067; UNIPARC:UPI0000000BE1; GB:M95603; NID:gi93875; PIDN:  
C:Keywords: DNA binding

Query Match 76.0%; Score 38; DB 2; Length 231;  
Best Local Similarity 83.3%; Pred. No. 99;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
Db 32 AAAAAAAAAA 43

#### RESULT 32

S11563  
probable MASH-2 protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S11563  
R:Johnson, J.E.; Birren, S.J.; Anderson, D.J.  
Nature 346, 858-861, 1990  
A:Title: Two rat homologues of Drosophila achaete-scute specifically expressed in neuro  
A:Reference number: S11562; MUID:90363294; PMID:2392153  
A:Accession: S11563  
A:Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-233 <JOH>





A;Note: sequence extracted from NCBI backbone (NCBIP:119832)

Query Match 76.0%; Score 38; DB 2; Length 331;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
||| |||||  
Db 83 AAAAAAAAAA 94

RESULT 40

A39065  
homeotic protein EVX2 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 31-Dec-2004  
C;Accession: A39065  
R;D'Esposito, M.; Morelli, F.; Acampora, D.; Migliaccio, E.; Simeone, A.; Boncinelli, E.;  
Genomics 10, 43-50, 1991  
A;Title: EVX2, a human homeobox gene homologous to the even-skipped segmentation gene,  
A;Reference number: A39065; MUID:91257849; PMID:1675198  
A;Accession: A39065  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-333 <DAE>  
A;Cross-references: UNIPARC:UPI000017A29D  
C;Genetics:  
A;Gene: GDB:EVX2  
A;Cross-references: GDB:127528; OMIM:142991  
A;Map position: 2q24.3-2q31  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;46-102/Domain: homeobox homology <HOX>

Query Match 76.0%; Score 38; DB 2; Length 333;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
||| |||||  
Db 213 AAAAAAAAAA 224

Search completed: September 9, 2006, 23:01:52  
Job time : 14.6709 secs

A;Residues: 1-323 <GER>  
A;Cross-references: UNIPARC:UPI000029F10; EMBL:X71422; NID:G397508; PIDN:CAA50553.1; PI  
R;Takada, S.; Cook, M.; Kramlauf, R.; McMahon, A.P.  
submitted to the EMBL Data Library, May 1991  
A;Description: Genomic sequence of mouse Hox-4.6.  
A;Reference number: S57443  
A;Accession: S57443  
A;Molecule type: DNA  
A;Residues: 'MNDCEGCPSSAAS', 1-323 <TRK>  
A;Cross-references: UNIPARC:UPI0000026A38; EMBL:X60395; NID:G871427; PIDN:CAA42943.1; PI  
C;Genetics:  
A;Gene: Hoxd-11  
A;Introns: 246/1  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;252-308/Domain: homeobox homology <HOX>

Query Match 76.0%; Score 38; DB 2; Length 323;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
||| |||||  
Db 93 AAAAAAAAAA 104

RESULT 38

AB1781  
hypothetical protein lin2792 [imported] - Listeria innocua (strain Clip11262)  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AB1781  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Krefit, J.; Kuhn, M.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AB1781  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-328 <GLA>  
A;Cross-references: UNIPROT:Q927J6; UNIPARC:UPI00000CC998; GB:AL592022; PIDN:CAC98018.1;  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin2792

Query Match 76.0%; Score 38; DB 2; Length 328;  
Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
||| |||||  
Db 313 AASASASASAA 324

RESULT 39

B47236  
zinc-finger protein Pur-1 - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 05-Apr-1995  
C;Accession: B47236  
R;Kennedy, G.C.; Rutter, W.J.  
Proc. Natl. Acad. Sci. U.S.A. 89, 11498-11502, 1992  
A;Title: Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates  
A;Reference number: A47236; MUID:93087555; PMID:1454839  
A;Accession: B47236  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-331 <KEN>  
A;Cross-references: UNIPARC:UPI000017C615  
A;Experimental source: insulinoma cell line T

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:38:21 ; Search time 109.975 Seconds  
(without alignments)  
100.934 Million cell updates/sec

Title: US-10-617-568-2

Perfect score: 50

Sequence: 1 AAAAAAAAAA 12

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : UniProt\_7.2.1\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	92.0	82	Q8VAT5	WSSV
2	44	88.0	112	Q3UUS5	MOUSE
3	44	88.0	233	Q7XTV6	ORYZA
4	44	88.0	309	Q3UK64	MOUSE
5	44	88.0	484	1 ZBTB8	MOUSE
6	44	88.0	484	2 Q3US18	MOUSE
7	44	88.0	998	2 Q3UJ04	MOUSE
8	43	86.0	235	2 Q8IGN6	DROME
9	43	86.0	395	2 Q6VRG5	OSTTA
10	43	86.0	475	2 Q3VEA3	DROME
11	42	84.0	305	2 Q8IMU7	DROME
12	42	84.0	637	2 Q70329	THETE
13	42	84.0	5233	1 HIW DROME	
14	41	82.0	85	2 Q9HPP9	HALSA
15	41	82.0	174	2 Q2XY87	DROER
16	41	82.0	243	2 Q3QK76	WHEAT
17	41	82.0	254	2 Q3U570	MOUSE
18	41	82.0	331	2 Q4B1F7	9BURK
19	41	82.0	375	1 SOX3	MOUSE
20	41	82.0	375	2 Q80XF1	MOUSE
21	41	82.0	403	2 Q6ODW4	ORYZA
22	41	82.0	417	2 Q75KP7	ORYZA
23	41	82.0	436	2 Q4VBD8	MOUSE
24	41	82.0	446	1 SOX3	HUMAN
25	41	82.0	446	2 Q5JWI3	HUMAN
26	41	82.0	448	2 Q8P564	MOUSE
27	41	82.0	449	2 Q3RKW0	MOUSE
28	41	82.0	515	2 Q6BC25	DROLT
29	41	82.0	524	2 Q8MQJ5	DROME
30	41	82.0	603	2 Q7F0I0	ORYZA
31	41	82.0	635	2 Q7PV80	ANOQA

#### ALIGNMENTS

32	41	82.0	642	2	Q9VE69	DROME
33	41	82.0	698	2	Q9NJZ3	DROME
34	41	82.0	705	2	Q7KSD2	DROME
35	41	82.0	1717	1	PHLPP	HUMAN
36	40	80.0	94	2	Q40629	ORYZA
37	40	80.0	109	2	Q5DFJ3	SCHJA
38	40	80.0	149	2	Q9A102	ORYZA
39	40	80.0	156	2	Q6ILH6	DROME
40	40	80.0	163	2	Q7YTA3	9MYRI
41	40	80.0	167	2	Q6Z8T9	ORYZA
42	40	80.0	168	2	Q2XYH2	DROYA
43	40	80.0	168	2	Q2XYH3	DROYA
44	40	80.0	178	2	Q33A26	ORYZA
45	40	80.0	216	1	RLI4	MOUSE
46	40	80.0	218	2	Q652J9	ORYZA
47	40	80.0	218	2	Q355G2	9BRAD
48	40	80.0	227	2	Q9CWK0	MOUSE
49	40	80.0	236	2	Q8GUP9	HUMAN
50	40	80.0	239	2	Q6CFV6	YARLI
51	40	80.0	241	2	Q8RYI1	ORYZA
52	40	80.0	245	2	Q7X6T2	ORYZA
53	40	80.0	245	2	Q6B6N8	DIDMA
54	40	80.0	253	2	Q8WZ46	HUMAN
55	40	80.0	268	2	Q4SBZ3	TETNG
56	40	80.0	269	2	Q6K3R9	ORYZA
57	40	80.0	303	2	Q3USA2	MOUSE
58	40	80.0	309	2	Q9NJY6	DROME
59	40	80.0	315	2	Q97601	RABIT
60	40	80.0	335	2	Q66V55	MAIZE
61	40	80.0	335	2	Q9FR09	MAIZE
62	40	80.0	349	2	Q7VU06	BORPE
63	40	80.0	349	2	Q7W473	BORPA
64	40	80.0	349	2	Q7WFN1	BORBR
65	40	80.0	352	2	Q6N082	HUMAN
66	40	80.0	361	2	Q4PEI8	USTMA
67	40	80.0	362	2	Q4SPTI	TETNG
68	40	80.0	368	2	Q7SBK2	NEUCR
69	40	80.0	376	1	EVE DROME	
70	40	80.0	377	1	HSF7	ARATH
71	40	80.0	377	2	Q98266	CERAE
72	40	80.0	378	1	HAIR	DROVI
73	40	80.0	378	2	Q9NAD6	HUMAN
74	40	80.0	378	2	Q81L66	DROER
75	40	80.0	392	1	HME1	HUMAN
76	40	80.0	392	2	Q4ZG44	HUMAN
77	40	80.0	395	2	Q80US0	MOUSE
78	40	80.0	395	2	Q8BKC1	MOUSE
79	40	80.0	395	2	Q3TSG4	MOUSE
80	40	80.0	395	2	Q8BKB9	MOUSE
81	40	80.0	401	1	HME1	MOUSE
82	40	80.0	401	2	Q3USF7	MOUSE
83	40	80.0	418	2	Q2QX18	ORYZA
84	40	80.0	428	1	FOX82	MOUSE
85	40	80.0	432	1	Q5VYV0	HUMAN
86	40	80.0	435	2	Q6GLG1	XENTR
87	40	80.0	435	2	Q2J548	9ACTO
88	40	80.0	441	2	Q6QD67	9PERC
89	40	80.0	448	2	Q7XVH0	ORYZA
90	40	80.0	458	2	Q6P6C2	HUMAN
91	40	80.0	459	1	RSP6	CHLRE
92	40	80.0	477	1	MAZ	HUMAN
93	40	80.0	477	2	Q97600	RABIT
94	40	80.0	490	2	Q9LJ04	ORYZA
95	40	80.0	493	2	Q8NFN7	HUMAN
96	40	80.0	506	2	Q3S2W8	ACHDO
97	40	80.0	508	2	Q7QGR6	ANOQA
98	40	80.0	509	2	Q76550	DROSI
99	40	80.0	509	2	Q9TWT5	DROSI
100	40	80.0	510	1	RUNT	DROME

Q9VE69	drosophila
Q9NJZ3	drosophila
Q7KSD2	drosophila
Q60346	h ph domain
Q40629	oryza sativ
Q5DFJ3	schistosoma
Q9A102	oryza sativ
Q6ILH6	drosophila
Q7YTA3	glomeris ma
Q6Z8T9	oryza sativ
Q2XYH2	drosophila
Q2XYH3	drosophila
Q33A26	oryza sativ
Q3CR57	mus musculus
Q652J9	oryza sativ
Q355G2	bradyrhizob
Q9CWK0	mus musculus
Q8GUP9	homo sapien
Q6CFV6	yarrowia li
Q8RYI1	oryza sativ
Q7X6T2	oryza sativ
Q6B6N8	didephis m
Q8WZ46	homo sapien
Q4SBZ3	tetraodon n
Q6K3R9	oryza sativ
Q3USA2	mus musculus
Q9NJY6	drosophila
Q97601	oryctolagus
Q66V55	zea mays (m
Q9FR09	zea mays (m
Q7VU06	bordetella
Q7W473	bordetella
Q7WFN1	bordetella
Q6N082	homo sapien
Q4PEI8	ustilago ma
Q4SPTI	tetraodon n
Q7SBK2	neurospora
P06602	drosophila
Q9T0D3	arabidopsis
Q98266	cercopithec
P29303	drosophila
Q9NAD6	homo sapien
Q81L66	drosophila
Q05925	homo sapien
Q4ZG44	homo sapien
Q80US0	mus musculus
Q8BKC1	m 0 day neo
Q3TSG4	mus musculus
Q8BKB9	mus musculus
P09065	mus musculus
Q3USF7	mus musculus
Q2QX18	oryza sativ
Q4733	mus musculus
Q5VYV0	homo sapien
Q6GLG1	xenopus tro
Q2J548	frankia sp.
Q6QD67	liparis atl
Q7XVH0	oryza sativ
Q6P6C2	homo sapien
Q1657	chlamydomon
P56270	homo sapien
Q97600	oryctolagus
Q9LJ04	oryza sativ
Q8NFN7	homo sapien
Q3S2W8	acheta dome
Q7QGR6	anopheles g
Q76550	drosophila
Q9TWT5	drosophila
P22814	drosophila

RESULT 1  
Q8VAT5\_WSSV PRELIMINARY; PRT; 82 AA.  
AC Q8VAT5;  
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2002, sequence version 1.  
DT 01-FEB-2006, entry version 12.  
DE WSV304 (WSSV360).  
OS White spot syndrome virus (WSSV).  
OC Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.  
OX NCBI\_TaxID=92652;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21548311; PubMed=11689662;  
RX DOI=10.1128/JVI.75.23.11811-11820.2001;  
RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.;  
RT "Complete genome sequence of the shrimp white spot bacilliform virus.";  
RL J. Virol. 75:11811-11820(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX MEDLINE=20517548; PubMed=11062040; DOI=10.1006/viro.2000.0597;  
RA Tsai M.F., Yu H.T., Tzeng H.F., Leu J.H., Chou C.M., Huang C.J.,  
RA Wang C.H., Lin J.Y., Kou G.H., Lo C.F.;  
RT "Identification and characterization of a shrimp white spot syndrome virus (WSSV) gene that encodes a novel chimeric polypeptide of cellular-type thymidine kinase and thymidylate kinase.";  
RL Virology 277:100-110(2000).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX MEDLINE=21844071; PubMed=11853398; DOI=10.1006/viro.2001.1273;  
RA Chen L.L., Leu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,  
RA Lo C.F., Kou G.H.;  
RT "Identification of a nucleocapsid protein (VP35) gene of shrimp white spot syndrome virus and characterization of the motif important for targeting VP35 to the nuclei of transfected insect cells.";  
RL Virology 293:44-53(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX MEDLINE=22192457; PubMed=12202227; DOI=10.1006/viro.2002.1480;  
RA Tzeng H.F., Chang Z.F., Peng S.E., Wang C.H., Lin J.Y., Kou G.H.,  
RA Lo C.F.;  
RT "Chimeric polypeptide of thymidine kinase and thymidylate kinase of shrimp white spot syndrome virus: thymidine kinase activity of the recombinant protein expressed in a baculovirus/insect cell system.";  
RL Virology 299:248-255(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX MEDLINE=22247910; PubMed=12359454; DOI=10.1006/viro.2002.1536;  
RA Chen L.L., Wang H.C., Huang C.J., Peng S.E., Chen Y.G., Lin S.J.,  
RA Chen W.Y., Dai C.F., Yu H.T., Wang C.H., Lo C.F., Kou G.H.;  
RT "Transcriptional analysis of the DNA polymerase gene of shrimp white spot syndrome virus.";  
RL Virology 301:136-147(2002).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX MEDLINE=22394561; PubMed=12504569; DOI=10.1006/viro.2002.1696;  
RA Lin S.F., Chang Y.S., Wang H.C., Tzeng H.F., Chang Z.F., Lin J.Y.,  
RA Wang C.H., Lo C.F., Kou G.H.;  
RT "Ribonucleotide reductase of shrimp white spot syndrome virus (WSSV): expression and enzymatic activity in a baculovirus/insect cell system and WSSV-infected shrimp.";  
RL Virology 304:282-290(2002).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX PubMed=15452257; DOI=10.1128/JVI.78.20.11360-11370.2004;

RA Tsai J.M., Wang H.C., Leu J.H., Hsiao H.H., Wang A.H., Kou G.H.,  
RA Lo C.F.;  
RT "Genomic and proteomic analysis of thirty-nine structural proteins of shrimp white spot syndrome virus.";  
RL J. Virol. 78:11360-11370(2004).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RX PubMed=15596810; DOI=10.1128/JVI.79.1.140-149.2005;  
RA Leu J.H., Tsai J.M., Wang H.C., Wang A.H., Wang C.H., Kou G.H.,  
RA Lo C.F.;  
RT "The unique stacked rings in the nucleocapsid of the white spot syndrome virus virion are formed by the major structural protein VP64, the largest viral structural protein ever found.";  
RL J. Virol. 79:140-149(2005).  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=Taiwan;  
RA Lo C.-F., Kou G.-H.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL; AF332093; AAL33306.1; -; Genomic DNA.  
DR EMBL; AF440570; AAL89228.1; -; Genomic DNA.  
SQ SEQUENCE 82 AA; 8635 MW; AF7E4C81AAF7D44 CRC64;  
Query Match 92.0%; Score 46; DB 2; Length 82;  
Best Local Similarity 91.7%; Pred. No. 35;  
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAMA 12  
Db 18 AAAAAAAAAAAMA 29  
||:|||||  
RESULT 2  
Q3UUS5\_MOUSE PRELIMINARY; PRT; 112 AA.  
AC Q3UUS5;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 11-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 5.  
DE 16 days neonate thymus cDNA, RIKEN full-length enriched library, clone: A130089M21 product: HBV pX associated protein 8 large isoform homolog.  
DE  
DE Name=Hbxap;  
GN Mus musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=C57BL/6J; TISSUE=Thymus;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M., Davis M.J., Walimig L.G., Aidinis V., Allen J.E., Ambesi-Impombato A., Apweiler R., Attaliya R.N., Bono H., Chalk A.M., Bansal M., Baxter L., Beisel K.W., Bersano T., Bono B., Della Gatta G., Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R., Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G., di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,

RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingersas T.R., Gojobori T., Green R.E.,  
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Humnietek L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jaki M., Kanapin A., Katoh M., Kawasawa Y., Kelson J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Motagui-Tabar S., Mulder N., Nakano N., Nakane K., Nakachi H., Ng P.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohata O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sanderlin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Sene S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugiyama K., Sultana R., Takenaka Y., Taki K.,  
RA Tannoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,  
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
RA Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,  
RA Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
RA Hayashizaki Y.;  
RT "The transcriptional landscape of the mammalian genome.";  
RL Science 309:1559-1563(2005).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group  
RT (Genome Network Core Team) and the FANTOM Consortium;  
RT "Antisense transcription in the Mammalian Transcriptome";  
RL Science 309:1564-1566(2005).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01286;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Kanagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltsev L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sanderlin A., Schneider C., Semple C.A., Setou M., Shmada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=21058660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK138087; BAE23549.1; -; mRNA.  
DR MGI; MGI:2682305; Hbvxap.  
SQ SEQUENCE 112 AA; 11657 MW; 3F364EA14DDFD076 CRC64;

Query Match 88.0%; Score 44; DB 2; Length 112;  
Best Local Similarity 91.7%; Pred. No. 83;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12  
|||  
DB 9 AAAAAAAAAA 20

RESULT 3

```
OQXTV6_ORYSA
ID QXTV6_ORYSA PRELIMINARY; PRT; 233 AA.
AC QXTV6;
DT 01-OCT-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2004, sequence version 2.
DT 07-FEB-2006, entry version 14.
DE OSJNB0010D21.10 protein.
DE OSJNB0010D21.10 protein.
GN Name=OSJNB0010D21.10;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartodeae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22337377; PubMed=12447439; DOI=10.1038/nature01183;
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y.,
RA Wang Q., Zhang L., Lu Y., Mu Y., Lu Y., Zhang L.S., Yu Z., Fan D.,
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,
RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu Q., Zhang X., Zhang W.,
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,
RA Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y.,
RA Han B.;
RT "Sequence and analysis of rice chromosome 4.";
RL Nature 420:316-320(2002).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL: AL606635; CAD41708.2; -; Genomic_DNA.
CC HSP: O80337.1GCC.
CC Gramene; QXTV6; -.
CC GO: GO:0005634; C:nucleus; IEA.
CC GO: GO:0003700; P:transcription factor activity; IEA.
CC GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
CC InterPro: IPR001471.1_TF_ERF.
CC Pfam: PF00847; AP2; 1.
CC PRINTS; PR00367; ETRHRSPELENT.
CC ProDom; PD001423; TF_ERF; 1.
CC SMART; SM00380; AP2; 1.
CC PROSITE; PS51032; AP2_ERF; 1.
CC SEQUENCE 233 AA; 25434 MW; 90B3BF38A3A4A419 CRC64;

Query Match 88.0%; Score 44; DB 2; Length 233;
Best Local Similarity 91.7%; Pred. NO. 1.5e+02;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAA 12
Db 18 AAAAAAAAAA 29

RESULT 4
Q3UK64_MOUSE
ID Q3UK64_MOUSE PRELIMINARY; PRT; 309 AA.
AC Q3UK64;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE T1B-55 BB88 cDNA, RIKEN full-length enriched library, clone:1730024D14
DE product:HBV pX associated protein 8 large isoform homolog (Fragment).
GN Name=HbXp;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RX STRAIN=BALB/C;
RA MEDLINE=223354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaudo I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami H., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
```

```

RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RX STRAIN=BALB/C;
RA PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodius R., Shimokawa K.,
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA Davis M.J., Wilming L.G., Aldinis V., Allen J.E.,
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA Di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA Hill D., Huminecki L., Iacono M., Ikeo K., Iwana A., Ishikawa T.,
RA Jakt M., Kanapin A., Katoh M., Kawasawa Y., Keiso J., Kitamura H.,
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P.,
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
RA Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K.,
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA Kawashima T., Kojima M., Kondo S., Konno K., Nakano K., Ninomiya N.,
RA Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S.,
RA Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RX STRAIN=BALB/C;
RA PubMed=16141073; DOI=10.1126/science.1112009;
RA RIKEN Genome Exploration Research Group, and Genome Science Group
RA (Genome Network Core Team) and the FANTOM Consortium;
RT "Antisense Transcription in the Mammalian Transcriptome.";
RL Science 309:1564-1566(2005).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RX STRAIN=BALB/C;
RA MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaudo I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami H., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
```

RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandler A., Schneider C., Sempie C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,  
RA Verardo R., Wagner L., Wylshew-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
RA Yaenueishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BALB/C;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli P., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BALB/C;  
RX MEDLINE=20493174; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RT prepare full-length cDNA libraries for rapid discovery of new genes";  
RL Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BALB/C;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RT sequencing pipeline with 384 multicapillary sequencer";  
RL Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=BALB/C;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.,  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.  
CC -----

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AK146154; BAE26938.1; -; mRNA.  
DR MGI; MGI:2682305; Hbxap.  
FT NON\_TER 309 309  
SQ SEQUENCE 309 AA; 34639 MW; 8099A6CAA99549E4 CRC64;  
Query Match 88.0%; Score 44; DB 2; Length 309;  
Best Local Similarity 91.7%; Pred. No. 1.9e+02;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAMA 12  
Db 7 AAAAAAAAAAAMA 18  
RESULT 5  
ZETB8 MOUSE  
ID ZETB8 MOUSE STANDARD; PRT; 484 AA.  
AC Q8C110;  
DT 26-APR-2004, integrated into UniProtKB/Swiss-Prot.  
DT 01-MAR-2003, sequence version 1.  
DT 07-MAR-2006, entry version 22.  
DE Zinc finger and BTB domain-containing protein 8.  
GN Name=Zbtb8;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muroidae; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].  
RC STRAIN=FVB/N;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywiniski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -!- FUNCTION: May be involved in transcriptional regulation.  
CC -!- SUBCELLULAR LOCATION: Nucleus (potential).  
CC -!- SIMILARITY: Contains 1 BTB (POZ) domain.  
CC -!- SIMILARITY: Contains 2 C2H2-type zinc fingers.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; BC023839; AAH23839.1; -; mRNA.  
DR Ensembl; ENSMUSG00000048485; Mus musculus.  
DR MGI; MGI:2387181; Zbtb8.  
DR InterPro; IPR000210; BTB.  
DR InterPro; IPR013069; BTB\_POZ.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf-C2H2; 2.  
DR ProDom; PD000003; Znf\_C2H2; 1.  
DR SMART; SM00225; BTB; 1.  
DR -----

DR SMART, SM00355; ZnF C2H2; 2.  
 DR PROSITE; PS00097; BTB; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
 KW DNA-binding; Metal-binding; Nuclear protein; Repeat; Transcription;  
 Transcription regulation; Zinc; Zinc-finger.  
 FT CHAIN 1 484  
 FT zinc finger and BTB domain-containing  
 FT protein 8.  
 FT /FTID=PRO\_0000047722.  
 FT BTB.  
 FT DOMAIN 24 92  
 FT ZN\_FING 331 353  
 FT ZN\_FING C2H2-type 1.  
 FT ZN\_FING C2H2-type 2.  
 FT COMPBTAS 132 149  
 FT Ala-rich.  
 FT SEQUENCE 484 AA; 53387 MW; B4FF082555BOCAIA CRC64;  
 Query Match 88.0%; Score 44; DB 1; Length 484;  
 Best Local Similarity 91.7%; Pred. No. 2.8e+02;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAA 12  
 Db 136 AAAAAAAAAA 147  
 RESULT 6  
 Q3US18\_MOUSE PRELIMINARY; PRT; 484 AA.  
 AC Q3US18;  
 DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
 DT 21-FEB-2006, entry version 7.  
 DE 16 days embryo head cDNA, RIKEN full-length enriched library,  
 DE clone:C130068L16 product:zinc finger and BTB domain containing 8, full  
 DE insert sequence (0 day neonate cerebellum cDNA, RIKEN full-length  
 DE enriched library, clone:C230068H07 product:zinc finger and BTB domain  
 DE containing 8, full insert sequence) (14, 17 days embryo head cDNA,  
 DE RIKEN full-length enriched library, clone:3222401820 product:zinc  
 DE finger and BTB domain containing 8, full insert sequence).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridea; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
 RX MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Methods Enzymol. 303:19-44(1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
 RX PubMed=16141072; DOI=10.1126/science.1112014;  
 RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
 Oyama R., Ravasi B., Lenhard B., Wells C., Kodzius R., Shimokawa K.,  
 Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,  
 Davis M.J., Wilming L.G., Aidinis V., Allen J.E., Ambesi-Impombato A.,  
 Apweiler R., Aurali R., Aurali R.N., Bailey T.L., Bansal M., Baxter L.,  
 Beisel K.W., Bersano T., Bono H., Chalk A.M., Chiu K.P., Choudhury V.,  
 Christoffels A., Clutterbuck D.R., Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
 di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,  
 Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
 Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E., Gustincich S.,  
 Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Humnietcki L.,  
 Iacono M., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
 Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V.,  
 Lareau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S.,  
 Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S.,  
 Miki H., Mignone F., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,

RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y.,  
 Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G., Petrovsky N.,  
 Piazza S., Reed J.F., Ring B.Z., Ringwald M., Rost B., Ruan Y.,  
 Salzberg S.L., Sandelin A., Schneider C., Schonbach C., Sekiguchi K.,  
 Semple C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada H.,  
 Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Suglura K.,  
 Sultana R., Takenaka Y., Taki K., Tammoja K., Tan S.L., Tang S.,  
 Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., van Nimwegen E.,  
 Verardo R., Wei C.L., Yagi K., Yamanishi H., Zabarovsky E., Zhu S.,  
 Zimmer A., Hide W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T.,  
 Brusic V., Quackenbush J., Wahlestedt C.,  
 Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S.,  
 Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J.,  
 Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T.,  
 Kojima M., Kondo S., Konno K., Nakano K., Ninomiya N., Nishio T.,  
 Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S., Tagami M.,  
 Waki K., Wataniki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
 RX MEDLINE=23546683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 Baldarelli R., Hill D.P., Bult C., Hume D.A., Batalov S., Beisel K.W.,  
 Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A.,  
 Bradt D., Brusic V., Chochia C., Corbani L.E., Cousins S., Dalla E.,  
 Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T.,  
 Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S.,  
 Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H.,  
 Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V.,  
 Lee Y., Lenhard B., Lyons P.H., Maglott D.R., Maltais L.,  
 Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K.,  
 Okido T., Pavan W.J., Pertea G., Petrovsky N., Pillai R., Pontius J.U.,  
 Qi D., Ramachandran S., Ravasi T., Reed J.C., Reid D.J., Ring B.Z., Ringwald M., Sandelin A.,  
 Schneider C., Semple C.A., Setou M., Shimada K., Sultana R.,  
 Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K., Verardo R.,  
 Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G.,  
 Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z.,  
 Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirokawa T.,  
 Kondo H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K.,  
 Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W.,  
 Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K.,  
 Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M.,  
 Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
 RX MEDLINE=21085660; PubMed=12117851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa K., Iizawa N., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casanova T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner T., Washio T.,



RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustinich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Natsuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaiji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=20493174; PubMed=11076861; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
RL prepare full-length cDNA libraries for rapid discovery of new genes.";  
RN Genome Res. 10:1617-1630(2000).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Head;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
RA Yanamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M.,  
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RT "RIKEN integrated sequence analysis (RISA) system-384-format  
RL sequencing pipeline with 384 multicapillary sequencer.";  
RN Genome Res. 10:1757-1771(2000).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Head;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
RA Hori F., Iida J., Inamura K., Imotani K., Itoh M., Kanagawa S.,  
RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AK140914; BA234519.1; -; mRNA.  
CC EMBL; AK163930; BA37541.1; -; mRNA.  
CC EMBL; AK132152; BA20999.1; -; mRNA.  
CC EMBL; GO:0005634; C:nucleus; IEA.  
CC GO; GO:0046872; F:metal ion binding; IEA.  
CC GO; GO:0003676; F:nucleic acid binding; IEA.  
CC GO; GO:0005515; F:protein binding; IEA.  
CC GO; GO:0008270; F:zinc ion binding; IEA.  
CC InterPro; IPR000210; BTB.  
CC InterPro; IPR013069; BTB\_POZ.  
CC Pfam; PF007087; Znf\_C2H2.  
CC Pfam; PF00651; BTB; 1.  
CC Pfam; PF00096; zf-C2H2; 2.  
CC ProDom; PD000003; Znf\_C2H2; 1.

DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; Znf\_C2H2; 2.  
DR PROSITE; PS00097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
KW Metal-binding; Nuclear\_protein; Repeat; Zinc; Zinc-finger.  
SQ SEQUENCE 484 AA; 53414 MW; ACA81DS402A2E7F4 CRC64;  
  
Query Match 88.0%; Score 44; DB 2; Length 484;  
Best Local Similarity 91.7%; Pred.No. 2.8e+02;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAA 12  
DB 136 AAAAAAAAAA 147  
  
RESULT 7  
Q3UU04 MOUSE  
ID Q3UU04 MOUSE PRELIMINARY; PRT; 998 AA.  
AC Q3UU04;  
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 07-FEB-2006, entry version 1.  
DE 17 days embryo stomach cDNA, RIKEN full-length enriched library,  
DE clone:1920040M16 product:hepatitis B virus x associated protein, full  
DE insert sequence. (Fragment).  
GN Name=Hbxap;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]\_TaxID=10090;  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Stomach;  
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RA Carninci P., Hayashizaki Y.;  
RT "High-efficiency full-length cDNA cloning.";  
RL Methods Enzymol. 303:19-44(1999).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Stomach;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,  
RA Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Zolotarev M.,  
RA Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Allen J.E.,  
RA Davis M.J., Wilming L.G., Aidinis V., Alizadeh A., Ambesi-Impombato A.,  
RA Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,  
RA Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,  
RA Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,  
RA Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,  
RA Di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,  
RA Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,  
RA Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,  
RA Guscinich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,  
RA Hill D., Humenick D., Iacono M., Ikeo K., Iwama A., Ishikawa T.,  
RA Jakt M., Kallian A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,  
RA Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,  
RA Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,  
RA Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,  
RA Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,  
RA Mottagui-Rabari S., Mulder N., Nakano N., Nakagawa S., Nori F., Ohara O.,  
RA Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,  
RA Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavoni G., Pesole G.,  
RA Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,  
RA Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,  
RA Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,  
RA Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,  
RA Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,  
RA Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,  
RA Ueda H.R., van Nimwegen B., Verardo R., Wei C.L., Yagi K.,  
RA Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,

RA Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,  
 RA Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,  
 RA Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,  
 RA Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,  
 RA Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,  
 RA Nishio T., Okada M., Plesky C., Shibata K., Shiraki T., Suzuki S.,  
 RA Tegami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,  
 RA Hayashizaki Y.;  
 RT "The transcriptional landscape of the mammalian genome.";  
 RL Science 309:1559-1563(2005).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG RIKEN Genome Exploration Research Group, and Genome Science Group  
 RG (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";  
 RL Science 309:1564-1566(2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Meglort D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Negashima T., Nunata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo L., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,  
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RX MEDLINE=20493374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [7]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RX MEDLINE=20530913; PubMed=11078661; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Stomach;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tagami M., Tegami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 CC EMBL; AK146675; BAE27351.1; -; mRNA.  
 DR MGI; MGI:2682305; Hbxap.  
 DR GO; GO:0005515; F:protein binding; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR001965; Znf\_PHD.  
 DR Pfam; PF00628; PHD; 1.  
 DR SMART; SM00249; PHD; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE; PS50016; ZF\_PHD\_2; 1.  
 DR NON\_TER 998 998  
 SQ SEQUENCE 998 AA; 110952 MW; 950D37671570EE0E CRC64.  
 Query Match 88.0%; Score 44; DB 2; Length 998;  
 Best Local Similarity 91.7%; Pred. No. 5e+02;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAAAMA 12  
 Db 9 AAAAAAAAAAAMA 20  
 |||||  
 |||||  
 RESULT 8  
 ID Q81GN6 DROME PRELIMINARY; PRT; 235 AA.  
 AC Q81GN6;  
 DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2003, sequence version 1.  
 DT 07-FEB-2006, entry version 8.  
 DE RE58281P.  
 DE ORFNames=CG18599;  
 OS Drosophila melanogaster (fruit fly).  
 GN Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;



RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
CC EMBL; AE003721; AAF5524.2; -; Genomic\_DNA.  
DR HSP; P06601; 1FJL.  
DR FlyBase; FBgn038592; CG18599.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0003700; P:transcription factor activity; IEA.  
DR GO; GO:0007275; P:development; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR013356; Homeobox.  
DR InterPro; IPR012287; Homeobox-rel.  
DR Pfam; PF00046; Homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW DNA-binding; Developmental protein; Homeobox; Nuclear protein;  
KW Transcription; transcription regulation.  
SQ SEQUENCE 475 AA; 50950 MW; 7DAC89752B1FA99D CRC64;  
  
Query Match 86.0%; Score 43; DB 2; Length 475;  
Best Local Similarity 83.3%; Pred. No. 3.7e+02;  
Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 AAAAAAAAAA 12  
Db 54 AAAAAAAAAA 65  
  
RESULT 11  
Q8IMU7 DROME Q8IMU7 DROME  
ID Q8IMU7 DROME PRELIMINARY; PRT; 305 AA.  
AC Q8IMU7;  
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2003, sequence version 1.  
DT 07-FEB-2006, entry version 14.  
DE CG31422-PA.  
GN ORFNames=CG31422, Dmel\_CG31422;

OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazell R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,

RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RG Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DDBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AE003750; AAN14029.1; -; Genomic\_DNA.  
DR FlyBase; FBgn0051422; CG31422.  
SQ SEQUENCE 305 AA; 32565 MW; 472F8AA40CF9953 CRC64;  
Query Match 84.0%; Score 42; DB 2; Length 305;  
Best Local Similarity 83.3%; Pred No. 3.5e+02;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
Db 271 AAAAAAAAAA 282  
RESULT 12  
ID Q70329 THETE PRELIMINARY; PRT; 637 AA.  
AC Q70329  
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.  
DT 05-JUL-2004, sequence version 1.  
DT 07-FEB-2006, entry version 9.  
DE Pyruvate-ferredoxin oxidoreductase and related 2-oxoacid-ferredoxin  
DE oxidoreductases gamma-alpha subunit (EC 1.-.-.-).  
GN Name-oor;  
OS Thermoproteus tenax.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoprotheales;  
OC Thermoproteaceae; Thermoproteus.  
OX NCBI\_TaxID=2271;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15028704; DOI=10.1128/JB.186.7.2179-2194.2004;  
RA Siebers B., Tjaden B., Michael K., Doerr C., Ahmed H., Zaparty M.,  
RA Gordon P., Senses C.W., Zibat A., Klenk H.-P., Schuster S.C.,  
RA Hensel R.;  
RT "Reconstruction of the central carbohydrate metabolism of  
RT Thermoproteus tenax using genomic and biochemical data.";  
RL J. Bacteriol. 186:2179-2194(2004).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AJ621314; CAF18495.1; -; Genomic DNA.  
DR GO; GO:0016491; P:oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002880; POR N.  
DR InterPro; IPR009014; Transketo\_C-like.  
DR Pfam; PF01855; POR\_N; 1.  
SQ Oxidoreductase; Pyruvate.  
SQ SEQUENCE 637 AA; 70670 MW; 2C5A0D9C4EA2AB5A CRC64;  
Query Match 84.0%; Score 42; DB 2; Length 637;

Best Local Similarity 83.3%; Pred. No. 6.4e+02;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
Db 308 AAAAAAAAAA 319  
RESULT 13  
HIW DROME  
ID HIW DROME STANDARD; PRT; 5233 AA.  
AC Q9NE7L; Q8MZ36; Q9VXZ5;  
DT 21-JUN-2005, integrated into UniProtKB/Swiss-Prot.  
DT 21-JUN-2005, sequence version 2.  
DT 07-MAR-2006, entry version 26.  
DE Ubiquitin ligase protein highwire (EC 6.3.2.-) (Protein  
DE pam/highwire/rpm-1).  
GN Name=hiw; ORFNames=CG32592;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, AND TISSUE SPECIFICITY.  
RC STRAIN=Canton-S;  
RX MEDLINE=20296162; PubMed=10839352; DOI=10.1016/S0896-6273(00)81166-6;  
RA Wan H.I., DiAntonio A., Fetter R.D., Bergstrom K., Strauss R.,  
RA Goodman C.S.;  
RT "Highwire regulates synaptic growth in Drosophila.";  
RL Neuron 26:313-329(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,  
Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
Hoskins N., Houghton K.A., Howland T.J., Wei M.-H., Ibegwack C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spiers R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S.-Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou J., Zhu S.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).

[3]  
 GENOME REANNOTATION.  
 RA MEDLINE=22426069; PubMed=12537572;  
 RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RA "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 [4]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 4083-5233.  
 RC STRAIN=Berkeley; TISSUE=Larva, and Pupae;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -!- FUNCTION: E3 ubiquitin ligase proteins mediate ubiquitination and  
 CC subsequent proteasomal degradation of target proteins. Has a role  
 CC during synaptogenesis; negative regulator of synaptic growth.  
 CC -!- PATHWAY: Ubiquitin conjugation; third step.  
 CC -!- SUBCELLULAR LOCATION: Nucleus (By similarity).  
 CC -!- TISSUE SPECIFICITY: Nervous system specific expression. Stage 13  
 CC embryos show expression in the CNS at the longitudinal axon tracts  
 CC around which the synaptic neuropil forms. Expression outside the  
 CC CNS starts at stage 16 in presynaptic terminals at the periaxial  
 CC zone which surround the active zone. NMJ and CNS expression is  
 CC also seen in third instar larvae (at protein level).  
 CC -!- MISCELLANEOUS: Flies lacking hwi display NMJ synapses that grow  
 CC exuberantly and are expanded in both the number of boutons and in  
 CC the extent and length of branches.  
 CC -!- SIMILARITY: Contains 1 filamin repeat.  
 CC -!- SIMILARITY: Contains 6 RCC1 repeats.  
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 -----  
 CC EMBL; AF626297; AAF76150.1; -; mRNA.  
 CC EMBL; AF003497; AAF48411.2; ALT\_SEQ; Genomic\_DNA.  
 CC EMBL; AY113376; AAM29381.1; ALT\_INIT; mRNA.  
 CC Ensembl; CG32592; Drosophila melanogaster.  
 CC FlyBase; FBgn0030600; hwi.  
 CC GO; GO:0016023; C:cytoplasmic membrane-bound vesicle; IDA.  
 CC GO; GO:0005886; C:plasma membrane; IDA.  
 CC GO; GO:0005515; F:protein binding; IPI.  
 CC GO; GO:0004842; F:ubiquitin-protein ligase activity; ISS.  
 CC GO; GO:0007628; P:adult walking behavior; NAS.  
 CC GO; GO:0030509; P:BMP signaling pathway; IPI.  
 CC GO; GO:0040011; P:locomotion; NAS.  
 CC GO; GO:0030514; P:negative regulation of BMP signaling pathway; IGI.  
 CC GO; GO:0045886; P:negative regulation of synaptic growth at n. .; IMP.  
 CC GO; GO:0016567; P:protein ubiquitination; ISS.  
 CC InterPro; IPR011989; ARM-like.  
 CC InterPro; IPR001298; Filamin.  
 CC InterPro; IPR012983; PHR.  
 CC InterPro; IPR000408; Reg\_chromatins.  
 CC InterPro; IPR001841; Znf\_RING.  
 CC Pfam; PF08005; PHR; 2.  
 CC Pfam; PF00415; RCC1; 1.  
 CC Pfam; PF00643; zf-B\_box; 1.  
 CC PRINTS; PR00633; RCNDNNGATON.  
 CC SMART; SM00184; RING; 1.  
 CC PROSITE; PS00194; FILAMIN\_REPEAT; 1.  
 CC PROSITE; PS00625; RCC1\_1; FALSE\_NEG.  
 CC PROSITE; PS00626; RCC1\_2; 2.

DR PROSITE; PS0012; RCC1\_3; 1.  
 DR PROSITE; PS00518; ZF\_RING\_1; FALSE\_NEG.  
 DR PROSITE; PS00089; ZF\_RING\_2; 1.  
 KW Complete proteome; Ligase; Metal-binding; Nuclear protein; Repeat;  
 KW Synapse; Transcription; Transcription regulation;  
 KW Ubiquitin conjugation pathway; Zinc; Zinc-finger.  
 FT CHAIN 1 5233  
 /FTID=PRO\_0000055971.  
 FT REPEAT 615 666  
 FT REPEAT 669 724  
 FT REPEAT 768 818  
 FT REPEAT 931 983  
 FT REPEAT 984 1033  
 FT REPEAT 1035 1084  
 FT REPEAT 1035 1084  
 FT REPEAT 2906 3000  
 FT ZN\_FING 4991 5042  
 FT CONFLICT 137 137  
 FT CONFLICT 604 604  
 FT CONFLICT 1620 1620  
 FT CONFLICT 4652 4652  
 FT CONFLICT 5233 AA; 565631 MW; E5220FFDC91A58E7 CRC64;  
 SQ SEQUENCE 5233 AA; 565631 MW; E5220FFDC91A58E7 CRC64;  
 Query Match 84.0%; Score 42; DB 1; Length 5233;  
 Best Local Similarity 90.9%; Pred. No. 3.6e+03;  
 Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAAAM 11  
 DB 45 AAAAAAAAAAAV 55  
 RESULT 14  
 Q9HPP9 HALSA PRELIMINARY; PRT; 85 AA.  
 AC Q9HPP9;  
 DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-MAR-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 16.  
 DE Vngl1530h.  
 GN OrderedLocusNames=VNG1530H;  
 OS Halobacterium salinarum (Halobacterium halobium).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=2242;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=NRC-1 / ATCC 700922 / JCM 11081;  
 RX MEDLINE=20504483; PubMed=11016950; DOI=10.1073/pnas.190337797;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weitz R., Goo Y.A.,  
 RA Leitchauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1.";   
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 -----  
 CC EMBL; AF005066; AAG19818.1; -; Genomic\_DNA.  
 DR PIR; F84306; F84306.  
 DR BioCyc; HSP64091.VNG1530H-MONOMER; -;  
 KW Complete proteome.  
 SQ SEQUENCE 85 AA; 8924 MW; 0CE3258BABC3F334 CRC64;  
 Query Match 82.0%; Score 41; DB 2; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAA 10

```

Db          30 AAAAAAAAAA 39
|||||
Q2XY87 DROER PRELIMINARY; PRT; 174 AA.
AC Q2XY87
DT 20-DEC-2005, integrated into UniProtKB/TrEMBL.
DT 20-DEC-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE CG13448 (Fragment).
OS Drosophila erecta (fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota;
OC Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=16120803; DOI=10.1093/molbev/msi246;
RA Cameron J.M., Guthrie T.B.;
RT "Intragenic Hill-Robertson Interference Influences Selection Intensity
  on Synonymous Mutations in Drosophila.";
RL Mol. Biol. Evol. 22:2519-2530(2005).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: DQ138775; ABA86381.1; -; Genomic_DNA.
FT NON_TER 174
SQ SEQUENCE 174 AA; 17861 MW; 76E5A0CDB39E7C7E CRC64;

Query Match 82.0%; Score 41; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 AAAAAAAAAA 10
|||||
Db          58 AAAAAAAAAA 67

RESULT 16
Q2QK76 WHEAT
ID Q2QK76 WHEAT PRELIMINARY; PRT; 243 AA.
AC Q2QK76
DT 24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT 24-JAN-2006, sequence version 1.
DT 07-FEB-2006, entry version 2.
DE Dehydration responsive element binding protein.
GN Name=DBP;
OS Triticum aestivum (Wheat).
OC Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BCP clade;
OC Poideae; Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Xu Z.S., Ma Y.Z., Cheng X.G., Li L.C., Chen M.;
RT "Isolation and functional identification of DRE binding protein in
  Triticum aestivum.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: DQ026517; AAY98505.1; -; mRNA.
SQ SEQUENCE 243 AA; 25932 MW; C2F8705208B73E9B CRC64;

Query Match 82.0%; Score 41; DB 2; Length 243;
Best Local Similarity 83.3%; Pred. No. 4e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy          1 AAAAAAAAAA 12
: |||||
Db          18 SAAAAAAAAA 29

RESULT 17
Q3U570 MOUSE
ID Q3U570 MOUSE PRELIMINARY; PRT; 254 AA.
AC Q3U570;
DT 11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT 11-OCT-2005, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched
  library, clone:C920021M21 product:muscleblind-like 1 (Drosophila),
  full insert sequence.
GN Name=Mbnl1;
OS Mus musculus (Mouse).
OC Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Methods Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX PubMed=16141072; DOI=10.1126/science.1112014;
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
  Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
  Bajic V.B., Bremner S.E., Batalov S., Forrest A.R., Zavolan M.,
  Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
  Ambesi-Impombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
  Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
  Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
  Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
  di Bernardo D., Down T., Engstrom P., Fagioli M., Faulkner G.,
  Fletcher C.F., Fukushima T., Furum M., Futaki S., Gariboldi M.,
  Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
  Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
  Hill D., Huminecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
  Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
  Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
  Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
  Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
  Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
  Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
  Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
  Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G.,
  Petrovsky N., Piazza S., Reed J.C., Reid J.F., Ring B.Z., Ringwald M.,
  Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
  Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
  Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
  Sperling S., Stupka E., Sugura K., Sultana R., Takenaka Y., Taki K.,
  Tammara K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
  Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
  Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
  Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
  Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki K., Tomaru Y.,
  Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
  Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
  Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
  Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
  Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J.,
  Hayashizaki Y.;
RT "The transcriptional landscape of the mammalian genome.";
RL Science 309:1559-1563(2005).
RN [3]
RP NUCLEOTIDE SEQUENCE.

```



RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX PubMed=16141073; DOI=10.1126/science.1112009;  
 RG Riken Genome Exploration Research Group, and Genome Science Group  
 (Genome Network Core Team) and the FANTOM Consortium;  
 RT "Antisense Transcription in the Mammalian Transcriptome.";   
 RL Science 309:1564-1566 (2005).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Buit C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Negashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Varardo R., Wagner L., Walestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
 RA Barney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";   
 RL Nature 420:563-573 (2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischnann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Buit C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";   
 RL Nature 409:685-690 (2001).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";   
 RL Genome Res. 10:1617-1630 (2000).  
 RN [7]

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Katsuna T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Orawa K., Tanaka T., Matsumura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";   
 RL Genome Res. 10:1757-1771 (2000).  
 RN [8]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=C57BL/6J; TISSUE=Thymus;  
 RA Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,  
 RA Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,  
 RA Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,  
 RA Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,  
 RA Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NonCommercial License  
 CC -----  
 DR EMBL; AK153848; BAE32210.1; -; mRNA.  
 DR MGI; MGI:1928482; Mbn11.  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:0005634; C:nucleus; IDA.  
 DR GO; GO:0006376; P:mRNA splice site selection; IMP.  
 DR InterPro; IPR000571; Znf\_CCH.  
 DR Pfam; PF00642; zf-CCH; 2.  
 DR SMART; SM00356; Znf\_C3H1; 1.  
 DR SEQUENCE 254 AA; -26605 MW; 3030142F792FAAE3 CRC64;  
 SQ  
 Query Match 82.0%; Score 41; DB 2; Length 254;  
 Best Local Similarity 83.3%; Pred. No. 41e+02;  
 Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAA 12  
 DB 166 AAQAAATAAAMA 177  
 RESULT 18  
 Q4B1F7\_9BURL PRELIMINARY; PRT; 331 AA.  
 AC Q4B1F7;  
 DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.  
 DT 13-SEP-2005, sequence version 1.  
 DT 21-FEB-2006, entry version 6.  
 DT Regulatory protein, LysR-LysR substrate binding domain precursor.  
 GN ORFNames=BprodRAFT\_3748;  
 OS Polaromonas sp. JS666.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Comamonadaceae; Polaromonas.  
 OX NCBI\_TaxID=296591;  
 RX [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JS666;  
 RG US DOE Joint Genome Institute (JGI-PGF);  
 RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,  
 RA Hammon N., Istrani S., Pitluck S., Richardson P.;  
 RA "Sequencing of the draft genome and assembly of Polaromonas sp.  
 RT JS666.";   
 RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JS666;  
 RG US DOE Joint Genome Institute (JGI-ORNL);  
 RA Larimer F., Land M.;



```
RT "Annotation of the draft assembly of Polaronomas sp. JS666.";
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=JS666;
RA US DOE Joint Genome Institute (JGI-PGF);
RG Copeland A., Lucas S., Lapidus A., Barry K., Dettler C., Glavina T.,
RA Hannon N., Israni S., Pittluck S., Richardson P.,
RA Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SIMILARITY: Contains 1 HTH LysR-type DNA-binding domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AF02000004; EAM39845.1; -; Genomic DNA.
CC GO; GO:0003700; P:transcription factor activity; IEA.
CC GO; GO:0006350; P:regulation of transcription; IEA.
CC GO; GO:0006350; P:transcription; IEA.
CC InterPro; IPR000847; HTH_LysR.
CC Pfam; PF00126; HTH_1; 1.
CC PRINTS; PR00039; HTH_LysR.
CC PROSITE; PS09331; HTH_LysR; 1.
CC DNA-binding; Signal; Transcription; Transcription regulation.
CC SIGNAL 309 331 Potential.
CC SEQUENCE 331 AA; 36159 MW; 2A4B7BA8EC46B78 CRC64;
SQ
Query Match 82.0%; Score 41; DB 2; Length 331;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAAAAAAAAA 10
Db 270 AAAAAAAAAA 279
RESULT 19
SOX3_MOUSE STANDARD; PRT; 375 AA.
ID _SOX3_MOUSE
AC P53784;
DT 01-OCT-1996, integrated into UniProtKB/Swiss-Prot.
DT 01-OCT-1996, sequence version 1.
DT 07-FEB-2006, entry version 35.
DE Transcription factor SOX-3.
GN Name=Sox3; Synonyms=Sox-3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=129/Sv;
RX MEDLINE=96189340; PubMed=8625802;
RA Collignon J., Sockanathan S., Hacker A., Cohen-Tannoudji M.,
RA Norris D., Rastan S., Stevanovic M., Goodfellow P.N., Lovell-Badge R.;
RA "A comparison of the properties of Sox-3 with Sry and two related
RA genes, Sox-1 and Sox-2.";
RT Development 122:509-520(1996).
RL [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=129/Sv;
RA Brunelli S., Bell D., Casey E.S., Hatland R., Lovell-Badge R.;
RA "Expression of Sox3 throughout the developing central nervous system
RA is dependent on the combined action of discrete, evolutionarily
RA conserved, regulatory elements.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nucleus.
CC -!- TISSUE SPECIFICITY: Mainly in the developing central nervous
CC system. Expressed in developing urogenital ridge.
CC -----
CC -!- SIMILARITY: Contains 1 HMG box DNA-binding domain.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AF434675; AAL40744.1; -; Genomic DNA.
CC HSSP; P48432; 1GTO.
CC SMR; P53784; 67-145.
CC Ensembl; ENSMUSG00000045179; Mus musculus.
CC MGI; MGI:98365; Sox3.
CC GO; GO:0030900; P:forebrain development; IMP.
CC GO; GO:0009887; P:organogenesis; IMP.
CC InterPro; IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 1.
CC SMART; SM00398; HMG; 1.
CC PROSITE; PS0118; HMG_BOX_2; 1.
CC DNA-binding; Nuclear protein; Transcription; Transcription regulation.
CC CHAIN 1 375 /FTID=PRO_0000048721.
FT DNA_BIND 69 137 HMG box.
FT COMPIAS 54 63 Poly-Gly.
FT COMPIAS 164 177 Poly-Ala.
FT COMPIAS 219 223 Poly-Pro.
FT COMPIAS 252 259 Poly-Ala.
FT COMPIAS 269 276 Poly-Ala.
FT COMPIAS 282 293 Poly-Ala.
SQ SEQUENCE 375 AA; 37858 MW; 2A8477A10A517FEF CRC64;
Query Match 82.0%; Score 41; DB 1; Length 375;
Best Local Similarity 83.3%; Pred. No. 5.7e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 AAAAAAAAAA 12
Db 284 AAAAAAAAAA 295
RESULT 20
SOX3_MOUSE PRELIMINARY; PRT; 375 AA.
ID _SOX3_MOUSE
AC Q80XFL1;
DT 01-JUN-2003, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 1.
DT 07-FEB-2006, entry version 17.
DE Transcription factor SOX3.
GN Name=Sox3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=129/Ola;
RX PubMed=15456859; DOI=10.1128/MCB.24.20.8834-8846.2004;
RA Tanaka S., Kamachi Y., Tanouchi A., Hamada H., Jing N., Kondoh H.;
RA "Interplay of SOX and POU factors in regulation of the Nestin Gene in
RA Neural Primordial Cells.";
RL Mol. Cell. Biol. 24:8834-8846(2004).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AB108674; BAC75669.1; -; mRNA.
CC HSSP; P48432; 1GTO.
CC SMR; Q80XFL1; 67-145.
CC Ensembl; ENSMUSG00000045179; Mus musculus.
CC MGI; MGI:98365; Sox3.
CC GO; GO:0005634; C:nucleus; RCA.
CC GO; GO:0003700; P:transcription factor activity; RCA.
CC GO; GO:0006325; P:establishment and/or maintenance of chromatin; RCA.
CC GO; GO:0030900; P:forebrain development; IMP.
```

```
DR GO; GO:0009887; P:organ morphogenesis; IMP
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; RCA.
DR GO; GO:0048515; P:spermatid differentiation; IMP.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS01118; HMG_BOX_2; 1.
SQ SEQUENCE 375 AA; 3779 MW; 3E117E20D40BF34 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 375;
Best Local Similarity 83.3%; Pred. No. 5.7e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAMAAAAAAMA 12
||:|||||
Db 284 AAAAAAAMA 295

RESULT 21
Q60DW4 ORYSA PRELIMINARY; PRT; 403 AA.
ID Q60DW4;
AC Q60DW4;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DE 07-FEB-2006, entry version 10.
DE Hypothetical protein P0426G01.8.
GN Names=P0426G01.8;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Chao Y.-T., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M.,
RA Chow T.-Y., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R.,
RA Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H.,
RA Hsiung J.-N., Heu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Leu H.-L.,
RA Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W.,
RA Wu H.-P., Shaw J.-F.;
RA "Oryza sativa PAC P0426G01 genomic sequence.";
RL Submitted (OCT-2004) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AC137623; AAV25643.1; -; Genomic_DNA.
Gramene; Q60DW4; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0006725; P:aromatic compound metabolism; IEA.
DR InterPro; IPR003089; AB hydrolase.
DR InterPro; IPR000073; AB_hydrolase_1.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00561; Abhydrolase_1; 1.
DR PRINTS; PR00111; ABHYDROLASE.
KW Hydrolase; Hypothetical protein.
SQ SEQUENCE 403 AA; 42774 MW; 30722980A480A479 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 403;
Best Local Similarity 83.3%; Pred. No. 6e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAMAAAAAAMA 12
||:|||||
Db 3 AAAAAAAMA 14

RESULT 22
Q75KP7 ORYSA PRELIMINARY; PRT; 417 AA.
ID Q75KP7;
AC Q75KP7;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 05-JUL-2004, sequence version 1.

21-FEB-2006, entry version 15.
Putative RING-H2 zinc finger protein.
Name=OSUNBa0065F09.2;
Oryza sativa (Japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
Ehrhartoideae; Oryzaceae; Oryza.
NCBI_TaxID=39947;
[1]
NUCLEOTIDE SEQUENCE.
Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
Fadrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
Vanaken S.S., Riedmuller S.B., Uterback T.T., Feldblyum T.V.,
Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
White O., Salzberg S.L., Fraser C.M.;
"Oryza sativa chromosome 3 BAC OSUNBa0065F09 genomic sequence.";
Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[2]
NUCLEOTIDE SEQUENCE.
Buell R.; (JAN-2004) to the EMBL/GenBank/DBJ databases.
Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
-!- DOMAIN: The RING-type zinc finger domain is essential for
ubiquitin ligase activity. It coordinates an additional third zinc
ion (by similarity).
-!- SIMILARITY: Contains 1 RING-type zinc finger.
-----
Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
Distributed under the Creative Commons Attribution-NoDerivs License
-----
EMBL; AC059323; AAR87274.1; -; Genomic_DNA.
Gramene; Q75KP7; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0008270; F;zinc ion binding; IEA.
DR GO; GO:0016567; P:protein ubiquitination; IEA.
DR InterPro; IPR001841; ZnF_RING.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.
SQ SEQUENCE 417 AA; 42736 MW; 7CC9F220D3B7883E CRC64;

Query Match 82.0%; Score 41; DB 2; Length 417;
Best Local Similarity 83.3%; Pred. No. 6.2e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAMAAAAAAMA 12
||:|||||
Db 101 AAAAAAAMA 112

RESULT 23
Q4VBD8 MOUSE PRELIMINARY; PRT; 436 AA.
ID Q4VBD8;
AC Q4VBD8;
DT 05-JUL-2005, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE Sox3 protein (fragment).
GN Name=Sox3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

```
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettunen M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RG NIH MGC Project;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC096018; AAH96018.1; -; mRNA.
DR SMR; Q4VBDB; 128-206.
DR MIM; MGI:98365; Sox3.
DR GO; GO:0005634; C:nucleus; RCA.
DR GO; GO:0003700; F:transcription factor activity; RCA.
DR GO; GO:0006325; P:establishment and/or maintenance of chromatin; RCA.
DR GO; GO:0006325; P:establishment and/or maintenance of chromatin; RCA.
DR GO; GO:0003090; P:forebrain development; IMP.
DR GO; GO:0009887; P:organ morphogenesis; IMP.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; RCA.
DR GO; GO:0048515; P:spermatid differentiation; IMP.
DR InterPro; IPR000910; HMG_BOX_2; 1.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS01118; HMG_BOX_2; 1.
FT NON TER 1
SQ SEQUENCE 436 AA; 43905 MW; C146DB68B042B6D1 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 436;
Best Local Similarity 83.3%; Pred. No. 6.4e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12
Db 345 AAAAAAAAAA 356

RESULT 24
SOX3 HUMAN
ID -SOX3 HUMAN STANDARD; PRT; 446 AA.
AC P41225; P35714; Q9NF49;
DT 01-FEB-1995, integrated into UniProtKB/Swiss-Prot.
DT 07-JUN-2004, sequence version 2.
DT 07-FEB-2006, entry version 38.
DE Transcription factor SOX-3.
GN Name=SOX3;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94154672; PubMed=8111369;
RA Stevanovic M., Lovell-Badge R., Collignon J., Goodfellow P.N.;
RT "SOX3 is an X-linked gene related to SRY.";
```

```
RL Hum. Mol. Genet. 2:2013-2018(1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RT Gorry M.C., Hart P.S., Sashi V., Hart T.C.;
RA "Clarification of the genomic sequence for human SOX3.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE OF 150-203.
RX MEDLINE=92310993; PubMed=1614875;
RA Denny P., Swift S., Brand N., Dabhadre N., Barton P., Ashworth A.;
RT "A conserved family of genes related to the testis determining gene,
RT SRY.";
RL Nucleic Acids Res. 20:2887-2887(1992).
CC -!- SUBCELLULAR LOCATION: Nucleus.
CC -!- SIMILARITY: Contains 1 HMG box DNA-binding domain.
CC -!- CAUTION: Was originally (Ref.3) termed SOX-9.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; X71135; CAA50465.1; -; Genomic DNA.
DR EMBL; AF264713; AAF73059.1; -; Genomic DNA.
DR EMBL; X65665; CAA46616.1; -; mRNA.
DR PIR; I38239; I38239.
DR PIR; S22942; S22942.
DR HSSP; P48432; LGT0.
DR SMR; P41225; 137-215.
DR TRANSFAC; T04916; -.
DR Ensembl; ENSG00000134595; Homo sapiens.
DR HGNC; HGNC:11199; SOX3.
DR MIM; 313430; gene.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0007417; P:central nervous system development; TAS.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS01118; HMG_BOX_2; 1.
KW DNA-binding; Nuclear protein; Transcription; Transcription regulation.
FT CHAIN 1 446
FT /FTID=PRO_0000048720.
FT DNA_BIND 139 207 HMG box.
FT COMEBIAS 129 133 Poly-Gly.
FT COMEBIAS 234 248 Poly-Ala.
FT COMEBIAS 290 294 Poly-Pro.
FT COMEBIAS 324 330 Poly-Ala.
FT COMEBIAS 340 347 Poly-Ala.
FT COMEBIAS 353 364 Poly-Ala.
FT CONFLICT 159 159 L -> Q (in Ref. 3).
FT CONFLICT 176 176 D -> E (in Ref. 3).
FT CONFLICT 202 202 E -> D (in Ref. 3).
FT CONFLICT 297 299 Missing (in Ref. 2).
SQ SEQUENCE 446 AA; 45210 MW; F79E79C2D16BB929 CRC64;

Query Match 82.0%; Score 41; DB 1; Length 446;
Best Local Similarity 83.3%; Pred. No. 6.5e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12
Db 355 AAAAAAAAAA 366

RESULT 25
Q5JWI3 HUMAN
ID Q5JWI3 HUMAN PRELIMINARY; PRT; 446 AA.
AC Q5JWI3;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DT 07-FEB-2006, entry version 6.
DE SRY (Sex determining region Y)-box 3.
GN Name=SOX3; ORFNames=RP11-51C14.3-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Way P.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RG NIH MGC Project;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AL121875; CAB87584.1; -; Genomic DNA.
CC EMBL; BC033865; AAH33865.1; -; mRNA.
CC EMBL; BC033863; AAH33863.1; -; mRNA.
CC SMR; Q5JW13; 137-215.
CC Ensembl; ENSG00000134595; Homo sapiens.
CC GO; GO:0003677; F:DNA binding; IEA.
CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
CC InterPro; IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 1.
CC SMART; SM00398; HMG; 1.
CC PROSITE; PS50118; HMG_BOX_2; 1.
CC SEQUENCE 446 AA; 45210 MW; F79E79C2D16BB929 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 446;
Best Local Similarity 83.3%; Pred. No. 6.5e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12
Db 355 AAAAAAAAAA 366

RESULT 26
Q6P564_MOUSE PRELIMINARY; PRT; 448 AA.
AC Q6P564;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Sox3 protein (Fragment).
GN Name=Sox3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; BC063061; AAH63061.1; -; mRNA.
CC SMR; Q6P564; 140-218.
CC Ensembl; ENSMUSG00000045179; Mus musculus.
CC MGI; MGI:98365; Sox3.
CC GO; GO:0005634; C:nucleus; RCA.
CC GO; GO:0003700; F:transcription factor activity; RCA.
CC GO; GO:0006325; P:establishment and/or maintenance of chromatin; RCA.
CC GO; GO:0003090; P:forebrain development; IMP.
CC GO; GO:0009887; P:organ morphogenesis; IMP.
CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; RCA.
CC GO; GO:0048515; P:spermatid differentiation; IMP.
CC InterPro; IPR000910; HMG_12_box.
CC Pfam; PF00505; HMG_box; 1.
CC SMART; SM00398; HMG; 1.
CC PROSITE; PS50118; HMG_BOX_2; 1.
CC NON_TER 1
CC SEQUENCE 448 AA; 45157 MW; BC567C0808CA64C9 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 448;
Best Local Similarity 83.3%; Pred. No. 6.6e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12
Db 357 AAAAAAAAAA 368

RESULT 27
Q5RKW0_MOUSE PRELIMINARY; PRT; 449 AA.
ID Q5RKW0;
AC Q5RKW0;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 10.
DE Sox3 protein (Fragment).
GN Name=Sox3;
OS Mus musculus (Mouse).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield V.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Director MGC Project;
RA Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC052024; AAHS2024.1; -; mRNA.
DR SMR; Q5RKW0; 141-219.
DR Ensembl; ENSMUSG00000045179; Mus musculus.
DR MGI; MGI:98365; Sox3.
DR GO; GO:0005634; C:nucleus; RCA.
DR GO; GO:0003700; P:transcription factor activity; RCA.
DR GO; GO:0006325; P:establishment and/or maintenance of chromatin; RCA.
DR GO; GO:0030900; P:forebrain development; IMP.
DR GO; GO:0009887; P:organ morphogenesis; IMP.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; RCA.
DR GO; GO:0048515; P:spermatid differentiation; IMP.
DR InterPro; IPR000910; HMG_12_box.
DR SMART; SM00398; HMG; 1.
DR PROSITE; PS00118; HMG_BOX_2; 1.
FT NON_TER 1
SQ SEQUENCE 449 AA; 45313 MW; B5386A3572C84DA7 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 449;
Best Local Similarity 83.3%; Pred. No. 6.6e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAAA 12
DB 358 AAAAAAAAAAAMS 369

RESULT 28
O6BCZ5 DROLT PRELIMINARY; PRT; 515 AA.
AC O6BCZ5;
DT 13-SEP-2004, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE CG4928 (Fragment).
OS Drosophila lutescens (Fruit fly).
```

```
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=51159;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15579698; DOI=10.1534/genetics.104.030478;
RA Swanson W.J., Wong A., Wolfner M.F., Aquadro C.F.;
RT "Evolutionary expressed sequence tag analysis of Drosophila female
RT reproductive tracts identifies genes subjected to positive
RT selection.";
RL Genetics 168:1457-1465(2004).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY665388; AAT76555.1; -; mRNA.
DR InterPro; IPR010291; DUF895_euk.
DR Pfam; PF05978; DUF895; 1.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 515 AA; 56186 MW; 77FBDDA7864EF077 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 515;
Best Local Similarity 83.3%; Pred. No. 7.4e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAAA 12
DB 491 AAAAAAAAAAALA 502

RESULT 29
O8MQJ5 DROME PRELIMINARY; PRT; 524 AA.
AC O8MQJ5;
DT 01-OCT-2002, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2002, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE LD22337p.
GN Name=CG7687; ORFNames=CG11122;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AY129447; AAM76189.1; -; mRNA.
DR FlyBase; FBgn0051122; CG11122.
SQ SEQUENCE 524 AA; 57365 MW; A43288FD17FC3DC0 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 524;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAAAAA 12
DB 163 AAQAAAAAANA 174
```

```
RESULT 30
Q7F010_ORYSA PRELIMINARY; PRT; 603 AA.
AC Q7F010;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-MAR-2006, entry version 16.
DE Hypothetical protein P0597G07.116.
GN Name=P0597G07.116;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC Ehrhartoidae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
RT clone:P0597G07.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AP004316; BAC83507.1; -; Genomic_DNA.
DR Gramene; Q7F010; -
DR GO; GO:0016021; C:membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019866; C:organelle inner membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0031072; F:heat shock protein binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR GO; GO:0006605; P:protein targeting; IEA.
DR GO; GO:0015031; P:protein transport; IEA.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR011990; TPR-like_helical.
DR InterPro; IPR013105; TPR_2.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF07719; TPR_2; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00076; DnaJ_2; 1.
DR Chaperone; Hypothetical protein; Inner membrane; Membrane;
KW Protein transport; Translocation; Transmembrane; Transport.
SQ SEQUENCE 603 AA; 64186 MW; 949D42CFD9143639 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 603;
Best Local Similarity 83.3%; Pred. NO. 8.4e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12
DB 508 AAAAAAAAAA 519

RESULT 31
Q7PV80_ANOGA PRELIMINARY; PRT; 635 AA.
AC Q7PV80;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 07-FEB-2004, sequence version 2.
DT 07-FEB-2006, entry version 8.
DE ENSANGP0000016650 (Fragment).
GN ORFNames=ENSANGG00000014161;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=PEST;
RC The Anopheles gambiae Sequence Committee;
RG
```

```
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AA001008986; EAA00301.3; -; Genomic_DNA.
DR InterPro; IPR004343; Plus-3.
DR Pfam; PF03126; Plus-3; 1.
DR SMART; SM00719; Plus3; 1.
DR NON_TER 1
FT SEQUENCE 635 AA; 71145 MW; 860CC8CFEC3D541C CRC64;

Query Match 82.0%; Score 41; DB 2; Length 635;
Best Local Similarity 83.3%; Pred. NO. 8.7e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAA 12
DB 98 AAAAAAAAAA 109

RESULT 32
Q9VE69_DROME PRELIMINARY; PRT; 642 AA.
AC Q9VE69;
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2003, sequence version 3.
DT 07-FEB-2006, entry version 25.
DE CG31122-PA.
GN ORFNames=CG31122, Dmel_CG31122;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiot E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
```



DE CG14307-PK, isoform K.  
GN Name:fru; ORFNames=Dmel\_CG14307;  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclele J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Paclele J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Paclele J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RX FlyBase;  
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AE003722; AAS65172.1; -; Genomic\_DNA.  
DR FlyBase; FBGN0004652; fru.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0005515; F:protein binding; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR InterPro; IPR000210; BTB.  
DR InterPro; IPR013069; Znf\_C2H2.  
DR Pfam; PF00651; BTB; 1.  
DR Pfam; PF00096; zf\_C2H2; 1.  
DR SMART; SM00225; BTB; 1.  
DR SMART; SM00355; Znf\_C2H2; 1.  
DR PROSITE; PS00097; BTB; 1.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 1.  
DR Metal-binding; Nuclear\_protein; Zinc; Zinc-finger.  
KW Metal-binding; Nuclear\_protein; Zinc; Zinc-finger.  
SQ SEQUENCE 705 AA; 76675 MW; F2D99E279741D9EE CRC64;  
Query Match 82.0%; Score 41; DB 2; Length 705;  
Best Local Similarity 100.0%; Pred. No. 9.5e+02;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAA 10  
Db 628 AAAAAAAAAA 637  
|||||  
RESULT 35  
PHLPP\_HUMAN  
ID PHLPP\_HUMAN STANDARD; PRT; 1717 AA.  
AC O60346; Q641Q7; Q6P4C4; Q6PTN6; Q96FK2; Q9NUY1;  
DC 07-JUN-2005, integrated into UniProtKB/Swiss-Prot.  
DT 07-JUN-2005, sequence version 3.  
DT 07-MAR-2006, entry version 41.  
DE PH domain leucine-rich repeat-containing protein phosphatase  
DE (EC 3.1.3.16) (PH domain leucine-rich repeat protein phosphatase)  
DE (pleckstrin homology domain-containing family E protein 1)  
DE (suprachiasmatic nucleus circadian oscillatory protein) (hSCOP).  
GN Name=PHLPP; Synonyms=KIAA0606; PLEKH1, SCOP;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;  
OC Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].



RA PubMed=16177791; DOI=10.1038/nature03983;  
RA Nussbaum C., Zody M.C., Borowsky M.B., Kamal M., Kodira C.D.,  
RA Taylor T.D., Whitaker C.A., Chang J.D., Cuomo C.A., Dewar K.,  
RA FitzGerald M.G., Yang X., Abouelleil A., Allen N.R., Anderson S.,  
RA Bloom T., Bugalter B., Butler J., Cook A., DeCaprio D., Engels R.,  
RA Garber M., Gnirke A., Hafez N., Hall J.L., Norman C.H., Itoh T.,  
RA Jaffe D.B., Kuroki Y., Lehoczy J., Lui A., Macdonald P., Mauceli E.,  
RA Mikelsen T.S., Naylor J.W., Nicol R., Nguyen C., Noguichi H.,  
RA O'Leary S.B., Pignani B., Smith C.L., Talamas J.A., Topham K.,  
RA Tokochi Y., Toyoda A., Wain H.M., Young S.K., Zeng Q., Zimmer A.R.,  
RA Fujiyama A., Hattori M., Birren B.W., Sakaki Y., Lander E.S.;  
RL "DNA sequence and analysis of human chromosome 18.";  
RL Nature 437:551-555(2005).  
[2]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 349-1717.  
RC TISSUE=Brain;  
RX MEDLINE=98290545; PubMed=9628581; DOI=10.1093/dnares/5.1.31;  
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,  
RA Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. IX.  
RT The complete sequences of 100 new cDNA clones from brain which can  
RT code for large proteins in vitro.";  
RL DNA Res. 5:31-39(1998).  
[3]  
RP SEQUENCE REVISION.  
RX MEDLINE=22158633; PubMed=12168954; DOI=10.1093/dnares/9.3.99;  
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
RT curation of 330 KIAA cDNA clones.";  
RL DNA Res. 9:99-106(2002).  
[4]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 491-1717.  
RC TISSUE=Brain, Eye, Lymph, and Skin;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
[5]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 495-1233.  
RC TISSUE=Placenta;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,  
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,

RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hisigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,  
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shitai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Iada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45(2004).  
[6]  
RP FUNCTION, ENZYME ACTIVITY, COFACTOR, ENZYME REGULATION, TISSUE  
RP SPECIFICITY, DISEASE, AND MUTAGENESIS OF 1715-THR--LEU-1717.  
RX PubMed=15808505; DOI=10.1016/j.molcel.2005.03.008;  
RA Gao T., Furnari F., Newton A.C.;  
RT "PHLPP: a phosphatase that directly dephosphorylates Akt, promotes  
RT apoptosis, and suppresses tumor growth.";  
RL Mol. Cell. 18:13-24(2005).  
CC -!- FUNCTION: Protein phosphatase that specifically mediates  
CC dephosphorylation of Ser-473 of AKT1, a protein that regulate the  
CC balance between cell survival and apoptosis through a cascade that  
CC primarily alters the function of transcription factors that  
CC regulate pro- and antiapoptotic genes. Dephosphorylation of Ser-  
CC 473 of AKT1 triggers apoptosis and suppression of tumor growth.  
CC May act as a negative regulator of K-Ras signaling in the membrane  
CC rafts.  
CC -!- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +  
CC phosphate.  
CC -!- COFACTOR: Binds 2 manganese ions per subunit (By similarity).  
CC -!- ENZYME REGULATION: Insensitive to okadaic acid.  
CC -!- SUBUNIT: Interacts with nucleotide free form K-Ras (KRAS) via its  
CC LRR repeats (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable). May be membrane-  
CC associated (By similarity).  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- DISEASE: Defects in PHLPP may be a cause of colon cancer. It is  
CC highly down-regulated in several colon cancer and glioblastoma  
CC cell lines with elevated AKT1 phosphorylation. Its reintroduction  
CC in such cell lines causes a strong suppression of tumor growth.  
CC -!- SIMILARITY: Contains 19 LRR (leucine-rich) repeats.  
CC -!- SIMILARITY: Contains 1 PH domain.  
CC -!- SIMILARITY: Contains 1 PP2C-like domain.  
CC -!- CAUTION: This is a conceptual translation.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL: AC015989; -; NOT ANNOTATED CDS; Genomic DNA.  
DR EMBL: AC022046; -; NOT ANNOTATED CDS; Genomic DNA.  
DR EMBL: AC027553; -; NOT ANNOTATED CDS; Genomic DNA.  
DR EMBL: AB011178; BAA25532.2; -; mRNA.  
DR EMBL: BC010706; AAH10706.1; -; mRNA.  
DR EMBL: BC014927; AAH14927.2; -; mRNA.  
DR EMBL: BC047653; AAH47653.1; -; mRNA.  
DR EMBL: BC063519; AAH63519.1; -; mRNA.  
DR EMBL: BC082244; AAH82244.1; ALT TERM; mRNA.  
DR EMBL: AK001924; BAA91980.1; ALT\_INT; mRNA.  
DR FIR: T00258; T00258.  
DR HSSP: P17778; 1JL5.  
DR Ensembl: ENSG00000081913; Homo sapiens.  
DR HGNC: HGNC:20610; PHLPP.  
DR MIM: 609396; gene.  
DR LinkHub: O60346; -.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003591; LRR\_typ.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR001932; PP2C-like.

DR Pfam; PF00560; LRR\_1; 19.  
 DR Pfam; PF00169; PH; 2.  
 DR Pfam; PF00481; PP2C; 2.  
 DR PRINTS; PR00019; LEURICHRPT.  
 DR SMART; SM00332; PP2Cc; 1.  
 DR PROSITE; PS00003; PH DOMAIN; 1.  
 KW Apoptosis; Hydrolase; Leucine-rich repeat; Manganese; Membrane;  
 KW Metal-binding; Protein phosphatase; Repeat.  
 KW PH domain leucine-rich repeat-containing  
 FT CHAIN 1 1717  
 FT FTId=PRO\_0000057781.  
 FT DOMAIN 536 636  
 FT REPEAT 636 659  
 FT REPEAT 661 686  
 FT REPEAT 690 713  
 FT REPEAT 715 736  
 FT REPEAT 737 759  
 FT REPEAT 761 784  
 FT REPEAT 806 829  
 FT REPEAT 830 853  
 FT REPEAT 871 893  
 FT REPEAT 894 916  
 FT REPEAT 917 939  
 FT REPEAT 941 961  
 FT REPEAT 962 984  
 FT REPEAT 986 1008  
 FT REPEAT 1011 1034  
 FT REPEAT 1035 1058  
 FT REPEAT 1059 1082  
 FT REPEAT 1104 1129  
 FT DOMAIN 1165 1420  
 FT REPEAT 1515 1538  
 FT MOTIF 1715 1717  
 FT COMPBIAS 4 7  
 FT COMPBIAS 21 45  
 FT COMPBIAS 132 137  
 FT COMPBIAS 138 144

Query Match 82.0%; Score 41; DB 1; Length 1717;  
 Best Local Similarity 83.3%; Pred. No. 2e+03;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAA 12  
 |||||  
 DB 31 AAAAAAAAAA 42

RESULT 36  
 Q0629 ORYSA  
 ID Q0629 ORYSA PRELIMINARY; PRT; 94 AA.  
 AC Q0629;  
 DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.  
 DT 01-NOV-1996, sequence version 1.  
 DT 07-FEB-2006, entry version 23.  
 DE Anther specific.  
 GN Name=RTS2;  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=IR54; TISSUE=Anther;  
 RA Lee J.-Y.K., Hodges T.K.;  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; U12171; AAA58584.1; -; Unassigned\_DNA.  
 DR PIR; T03285; T03285.  
 DR Gramene; Q0629; -.

SQ SEQUENCE 94 AA; 8628 MW; DAD82B0F05ED1B8F CRC64;  
 Query Match 80.0%; Score 40; DB 2; Length 94;  
 Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
 Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AAAAAAAAAA 12  
 |||||  
 DB 14 AAAAAAAAAA 24

RESULT 37  
 Q5DFJ3 SCHJA  
 ID Q5DFJ3 SCHJA PRELIMINARY; PRT; 109 AA.  
 AC Q5DFJ3;  
 DT 29-MAR-2005, integrated into UniProtKB/TrEMBL.  
 DT 29-MAR-2005, sequence version 1.  
 DT 07-FEB-2006, entry version 4.  
 DE SJCHG07588 protein.  
 OS Schistosoma japonicum (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 OX NCBI\_TaxID=6182;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Han Z.;  
 RT "The full-length cDNA sequences of Schistosoma japonicum genes.";  
 RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Liu F., Lu J., Hu W., Wang S.-Y., Cui S.-J., Chi M., Yan Q.,  
 Wang X.-R., Song H.-D., Xu X.-N., Wang J.-J., Zhang X.-L., Wang Z.-Q.,  
 Xue C.-L., Brindley P.J., McManus D.P., Yang P.-Y., Feng Z., Chen Z.,  
 Han Z.-G.;  
 RT "New Perspectives on Host-parasite Interplay by Comparative  
 RT Transcriptomic and Proteomic Analyses of the Human Blood Fluke,  
 RT Schistosoma japonicum.";  
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; AY813681; AAW25413.1; -; mRNA.  
 SQ SEQUENCE 109 AA; 10655 MW; 2EA7EF32B06A6CEP CRC64;  
 Query Match 80.0%; Score 40; DB 2; Length 109;  
 Best Local Similarity 83.3%; Pred. No. 2.8e+02;  
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAA 12  
 |||||  
 DB 34 AAAAAAAAAA 45

RESULT 38  
 Q94102 ORYSA  
 ID Q94102 ORYSA PRELIMINARY; PRT; 149 AA.  
 AC Q94102;  
 DT 01-DEC-2001, integrated into UniProtKB/TrEMBL.  
 DT 01-DEC-2001, sequence version 1.  
 DT 07-FEB-2006, entry version 11.  
 DE Hypothetical protein OSUNBa0034E23.23.  
 GN Name=OSUNBa0034E23.23;  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Nipponbare;  
 RA McCombie W.R.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

```
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AC022352; AAK51587.1; -; Genomic_DNA.
KW Gramine; Q94T02; -.
KW Hypothetical protein.
SQ SEQUENCE 149 AA; 14809 MW; E9125D3203F934A2 CRC64;

Query Match      80.0%; Score 40; DB 2; Length 149;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12
Db 60 AALAAAAAVALA 71

RESULT 39
Q6ILH6 DROME PRELIMINARY; PRT; 156 AA.
AC Q6ILH6;
DR 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE HDC09438.
GN ORFNames=HDC09438;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14709175; DOI=10.1186/gb-2003-5-1-r3;
RA Hild M., Beckmann B., Haas S.A., Koch B., Solovyev V., Busold C.,
RA Fellenberg K., Boutros M., Vingron M., Sauer F., Hoheisel J.D.,
RA Paro R.;
RT "An integrated gene annotation and transcriptional profiling approach
RT towards the full gene content of the Drosophila genome.";
RL Genome Biol. 5:RESEARCH0003.1-RESEARCH0003.17(2003).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BK002040; DAA02885.1; -; Genomic DNA.
SQ SEQUENCE 156 AA; 16458 MW; 2EA5CBB3370E61C CRC64;

Query Match      80.0%; Score 40; DB 2; Length 156;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12
Db 52 AAAAAAAAAA 63

RESULT 40
Q7YTA3 9MYRI PRELIMINARY; PRT; 163 AA.
AC Q7YTA3;
DR 01-OCT-2003, integrated into UniProtKB/TrEMBL.
DT 01-OCT-2003, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Distal-less protein (Fragment).
GN Name=dll;
OS Glomeris marginata.
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Diplopoda; Pentazonia;
OC Glomerida; Glomeridae; Glomeris.
OX NCBI_TaxID=62006;
RN 1]
```

```
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22767835; PubMed=12885558; DOI=10.1016/S0012-1606(03)00217-3;
RA Eric N.M., Tautz D.;
RT "The expression of the proximodistal axis patterning genes Distal-less
RT and dach shund in the appendages of Glomeris marginata (Myriapoda:
RT Diplopoda) suggests a special role for these genes in patterning the
RT head appendages.";
RL Dev. Biol. 260:97-112(2003).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AJ551276; CAD82905.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR012287; Homeodomain-rel.
DR Pfam; PF00046; Homeobox; 1.
DR ProDom; PD000010; Homeobox; 1.
DR PROSITE; PS00071; HOMEBOX 2; 1.
KW DNA-binding; Homeobox; Nuclear protein.
FT NON_TER 1 163
FT NON_TER 163 163
SQ SEQUENCE 163 AA; 17588 MW; 87B8BCD9F5421697 CRC64;

Query Match      80.0%; Score 40; DB 2; Length 163;
Best Local Similarity 90.9%; Pred. No. 3.9e+02;
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 11
Db 6 AAAAAAAAAA 16

Search completed: September 9, 2006, 23:00:26
Job time : 113.975 secs
```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:37:17 ; Search time 97.4177 Seconds  
(without alignments)  
61.014 Million cell updates/sec

Title: US-10-617-568-3

Perfect score: 53

Sequence: 1 AAAAAAAAAAAAAA 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A\_Geneseq\_8.\*

1: geneseqp1980s.\*

2: geneseqp1990s.\*

3: geneseqp2000s.\*

4: geneseqp2001s.\*

5: geneseqp2002s.\*

6: geneseqp2003as.\*

7: geneseqp2003bs.\*

8: geneseqp2004s.\*

9: geneseqp2005s.\*

10: geneseqp2006s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	13	ADI29007	ADI29007 Human MHC
2	53	100.0	278	ADR98917	ADR98917 Lung spec
3	53	100.0	646	ADY09213	ADY09213 Plant ful
4	50	94.3	333	ADX78446	ADX78446 Plant ful
5	49	92.5	109	ABO58367	ABO58367 Human gen
6	49	92.5	112	ABG15501	ABG15501 Novel hum
7	49	92.5	252	AAQ16493	AAQ16493 Zorocrate
8	49	92.5	252	AAE36894	AAE36894 Zorocrate
9	49	92.5	285	ABM65204	ABM65204 Drosophil
10	49	92.5	316	ABG18917	ABG18917 Novel hum
11	49	92.5	337	ABM69577	ABM69577 Drosophil
12	49	92.5	373	AAE35778	AAE35778 Human SEC
13	49	92.5	377	ADG61835	ADG61835 Transcrip
14	49	92.5	378	ABP65077	ABP65077 Hypoxia-1
15	49	92.5	393	AAAM25623	AAAM25623 Human pro
16	49	92.5	394	AAE35774	AAE35774 Human SEC
17	49	92.5	512	ADB64824	ADB64824 Human pro
18	49	92.5	578	ABM59750	ABM59750 Drosophil
19	49	92.5	708	AAE92956	AAE92956 Human pro
20	49	92.5	815	ABB70499	ABB70499 Drosophil
21	49	92.5	815	ABP54340	ABP54340 Drosophil
22	49	92.5	855	ABB71577	ABB71577 Drosophil
23	49	92.5	892	ABU64278	ABU64278 Murine ub

24	49	92.5	892	7	ADH73564	ADH73564 Mouse POS
25	49	92.5	892	8	ADR89697	ADR89697 Murine PO
26	49	92.5	892	8	ADR34245	ADR34245 POSH prot
27	49	92.5	892	8	ADT77904	ADT77904 Mouse POS
28	49	92.5	892	8	ADU73761	ADU73761 Mouse POS
29	49	92.5	892	8	ADU68997	ADU68997 House mou
30	49	92.5	892	9	ADW87412	ADW87412 Mouse POS
31	49	92.5	892	9	ADZ66429	ADZ66429 Mouse POS
32	49	92.5	892	9	AEA10631	AEA10631 Murine PO
33	49	92.5	892	9	AEC94736	AEC94736 Mouse POS
34	49	92.5	892	4	ABG05909	ABG05909 Novel hum
35	49	92.5	892	4	ABM66003	ABM66003 Drosophil
36	49	92.5	1235	4	ABM58493	ABM58493 Drosophil
37	49	92.5	1444	4	ABG15667	ABG15667 Novel hum
38	49	92.5	1937	4	ABM58985	ABM58985 Drosophil
39	49	92.5	1944	6	AAE36107	AAE36107 Human chr
40	49	92.5	1944	7	ADI26773	ADI26773 Human chr
41	49	92.5	1944	7	ADI26772	ADI26772 Human CHD
42	49	92.5	1944	7	ADJ68739	ADJ68739 Human hea
43	49	92.5	2000	7	ADI26747	ADI26747 Human GS-
44	48	90.6	13	8	ADI29008	ADI29008 Human MHC
45	47	88.7	13	9	ADV78640	ADV78640 Cell atta
46	47	88.7	13	10	AED86538	AED86538 Silk pept
47	47	88.7	14	9	ADV78641	ADV78641 Cell atta
48	47	88.7	14	10	AED86539	AED86539 Silk pept
49	47	88.7	15	5	ABP52215	ABP52215 HLA-DR2 m
50	47	88.7	15	6	ABR38361	ABR38361 Human can
51	47	88.7	15	6	ABR38363	ABR38363 Human can
52	47	88.7	15	6	ABR38360	ABR38360 Human can
53	47	88.7	15	6	ABR38364	ABR38364 Human can
54	47	88.7	15	6	ABR38365	ABR38365 Human can
55	47	88.7	15	6	ABR38362	ABR38362 Human can
56	47	88.7	15	9	ADO58104	ADO58104 Spider be
57	47	88.7	15	9	ADV78642	ADV78642 Cell atta
58	47	88.7	15	10	AED86540	AED86540 Silk pept
59	47	88.7	15	10	AEF42853	AEF42853 Arg-C pro
60	47	88.7	16	2	AAW32224	AAW32224 Beta-shee
61	47	88.7	16	2	AAW32226	AAW32226 Beta-shee
62	47	88.7	16	9	ADV78643	ADV78643 Cell atta
63	47	88.7	17	2	AAW32227	AAW32227 Beta-shee
64	47	88.7	18	7	AAO30390	AAO30390 Human ARX
65	47	88.7	18	8	ADH58940	ADH58940 Silk prot
66	47	88.7	19	10	AEF40057	AEF40057 Polyalan
67	47	88.7	19	7	AEO30392	AEO30392 Human ARX
68	47	88.7	19	9	ADV78587	ADV78587 Cell atta
69	47	88.7	19	9	AEB23102	AEB23102 Cell adhe
70	47	88.7	20	4	AAE69489	AAE69489 Synthetic
71	47	88.7	20	7	ADH47641	ADH47641 Human lun
72	47	88.7	20	8	ADH58955	ADH58955 Silk prot
73	47	88.7	20	9	ADJ21560	ADJ21560 Human lun
74	47	88.7	20	9	ADV78588	ADV78588 Cell atta
75	47	88.7	20	10	AEF40072	AEF40072 Domestic
76	47	88.7	21	4	AAE69490	AAE69490 Synthetic
77	47	88.7	21	7	ADJ81212	ADJ81212 Self-asse
78	47	88.7	21	7	ADJ81342	ADJ81342 Self-asse
79	47	88.7	21	9	ADV78589	ADV78589 Cell atta
80	47	88.7	22	9	ADV78590	ADV78590 Cell atta
81	47	88.7	22	9	ADV78591	ADV78591 Cell atta
82	47	88.7	23	9	ADV78592	ADV78592 Cell atta
83	47	88.7	25	5	ABG95821	ABG95821 Transmemb
84	47	88.7	26	7	ADJ81343	ADJ81343 Self-asse
85	47	88.7	26	7	ADJ81213	ADJ81213 Self-asse
86	47	88.7	26	9	AEB23106	AEB23106 Cell adhe
87	47	88.7	27	10	AEF42855	AEF42855 Asp-N end
88	47	88.7	28	7	ADJ81344	ADJ81344 Self-asse
89	47	88.7	28	7	ADJ81214	ADJ81214 Self-asse
90	47	88.7	28	9	AEB23104	AEB23104 Cell adhe
91	47	88.7	28	9	AEB23105	AEB23105 Cell adhe
92	47	88.7	31	3	AAE08166	AAE08166 Peptide m
93	47	88.7	32	9	ADH83853	ADH83853 Heparin b
94	47	88.7	38	9	ADV78600	ADV78600 Cell atta
95	47	88.7	39	5	ABG95847	ABG95847 Fusion pe
96	47	88.7	40	9	ADV78601	ADV78601 Cell atta

97 47 88.7 40 9 AEB23103 Aeb23103 Cell adhe  
98 47 88.7 40 9 AED01242 Aed01242 Wound dre  
99 47 88.7 40 9 AEE05648 Aee05648 Auxilliary  
100 47 88.7 40 9 AEE01614 Aee01614 Auxilliary

## ALIGNMENTS

RESULT 1  
ADI29007  
ID ADI29007 standard; peptide; 13 AA.  
XX AC  
AC ADI29007;  
DT  
DT 15-APR-2004 (first entry)  
XX  
DE Human MHC class II compound spaceholder molecule SEQ ID NO:3.  
XX  
XX MHC class II compound; MHC class II component; MHC class II alpha chain;  
KW MHC class II beta chain; peptide binding groove; spaceholder molecule;  
KW effector component; immune response; immune disorder; virucide;  
KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
KW gene therapy; viral infections; bacterial infection; parasitic infection;  
KW neoplastic disease; autoimmunity; toxicity; human.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX WO2004007528-A2.  
XX  
XX 22-JAN-2004.  
XX  
XX 11-JUL-2003; 2003WO-US021767.  
XX  
XX 12-JUL-2002; 2002US-0395494P.  
XX  
XX 22-JUL-2002; 2002US-0397893P.  
XX  
XX (DAND ) DANA FARBER CANCER INST INC.  
XX  
XX Wucherpfenning KW, Seth N;  
XX  
XX WPI; 2004-122876/12.  
XX  
XX New MHC class II compound, useful for preparing a composition for  
PT treating immune disorders e.g. viral infections, bacterial infections,  
PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
XX  
XX Claim 11; SEQ ID NO 3; 92pp; English.  
XX  
XX The present invention describes an isolated MHC class II compound (I)  
CC comprising: (a) an MHC class II component comprising at least a portion  
CC of an MHC class II alpha chain and at least a portion of an MHC class II  
CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
CC chain form a peptide binding groove; (b) a spaceholder molecule; and (c)  
CC an effector component, where the effector component is linked to the MHC  
CC class II component. Also described: (1) a pharmaceutical composition  
CC comprising the MHC class II molecule and a carrier; (2) a method of  
CC producing an MHC class II compound; (3) a method of directly identifying  
CC an antigen-specific T cell; (4) a method of regulating an immune response  
CC in a subject; (5) a method of treating an immune disorder in a subject;  
CC and (6) a method of regulating an immune response ex vivo in a subject; and  
CC (7) a method of treating an immune disorder ex vivo in a subject. (1) has  
CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive  
CC activities, and can be used in gene therapy. The MHC class II compound  
CC (I) can be used for preparing a composition for treating immune  
CC disorders, e.g., viral infections, bacterial infections, parasitic  
CC infections, neoplastic disease, autoimmunity or toxicity. The present  
CC sequence represents a spaceholder molecule peptide, which can be used in  
CC an MHC class II compound from the present invention.

XX Sequence 13 AA;

Query Match 100.0%; Score 53; DB 8; Length 13;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
| | | | | | | | | |  
Db 1 AAAAAAAAAAAAA 13

## RESULT 2

ADR98917  
ID ADR98917 standard; protein; 278 AA.  
XX  
AC ADR98917;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Lung specific gene splice variant encoded protein #89.  
XX  
KW cytostatic; gene therapy; vaccine; lung; diagnosis; cancer;  
KW non-cancerous lung disease; lung tissue; antagonist; gene therapy;  
KW transgenic animal; splice variant.  
XX  
OS Homo sapiens.  
XX WO2004074430-A2.  
XX  
XX 02-SEP-2004.  
XX  
XX 08-DEC-2003; 2003WO-US038896.  
XX  
XX 06-DEC-2002; 2002US-0431307P.  
XX  
XX 06-DEC-2002; 2002US-0431510P.  
XX  
XX 06-DEC-2002; 2002US-0431516P.  
XX  
XX (DIAD-) DIADEXUS INC.  
XX  
XX Macina RA, Turner LR, Sun Y, Liu S;  
XX  
XX WPI; 2004-635553/61.  
XX  
XX N-PSDB; ADR98780.  
XX  
XX New isolated human lung specific nucleic acid molecule, useful for  
PT identifying, diagnosing, monitoring, staging, imaging and treating lung  
PT cancer and non-cancerous diseases of the lung.  
XX  
XX Claim 1; SEQ ID NO 202; 542pp; English.  
XX  
XX The invention relates to a new isolated lung specific nucleic acid  
CC molecule (I) comprising any of 113 fully defined nucleotide sequences  
CC given in the specification, their encoded protein sequences, sequences  
CC selectively hybridizing to the nucleotide sequences or a sequence having  
CC at least 60% identity to the nucleotide sequences. The methods and  
CC compositions of the present invention are useful for identifying,  
CC diagnosing, monitoring, staging, imaging and treating lung cancer and non  
CC -cancerous diseases of the lung. They are also used for identifying lung  
CC tissue, monitoring and identifying and/or designing antagonists of the  
CC polypeptide of the invention, gene therapy, production of transgenic  
CC animals and production of engineered lung tissue for treatment and  
CC research. Lung specific genes (LSGs) were identified by a systematic  
CC analysis of gene expression data in the LIFESEQ Gold database using the  
CC data mining software package candidate lead automatic search program  
CC (CLASP). Genes were grouped into gene bins where each bin is a cluster of  
CC sequences grouped together where they share a common contig.  
CC Differentially expressed tissue-specific genes were selected based on the  
CC percentage level in the targeted tissue versus all the other tissues. The  
CC expression levels for each gene in libraries of normal tissues or non-  
CC tumour tissues from cancer patients were compared with the expression  
CC levels in tissue libraries associated with tumour or disease. This  
CC sequence represents a protein of the invention.

XX Sequence 278 AA;

Query Match 100.0%; Score 53; DB 8; Length 278;  
Best Local Similarity 100.0%; Pred. No. 4.2;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
| | | | | | | | | |  
Db 1 AAAAAAAAAAAAAA 13

RESULT 3  
ID ADY09213 standard; protein; 646 AA.  
AC ADY09213;  
XX  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polypeptide seqid 65028.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content.  
XX  
OS Unidentified.  
XX  
XX US2004034888-A1.  
XX  
XX 19-FEB-2004.  
XX  
XX  
XX 28-APR-2003; 2003US-00425114.  
XX  
XX 06-MAY-1999; 99US-00304517.  
XX  
XX 05-NOV-2001; 2001US-00985678.  
XX  
XX (LIUJ/) LIU J.  
XX (ZHOU/) ZHOU Y.  
XX (KOVA/) KOVALIC D K.  
XX (SCRE/) SCREEN S E.  
XX (TABA/) TABASKA J E.  
XX (CAOY/) CAO Y.  
XX  
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
XX pests, for conferring increased resistance to plant disease, or for  
XX improving yield.  
XX  
XX Claim 1; SEQ ID NO 65028; 15pp; English.

The invention describes a recombinant DNA construct comprising a polynucleotide consisting of a sequence encoding an amino acid sequence available in electronic form from the US patent office at ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide of the invention are also useful in physical arrays of molecules and as plant breeding markers. The recombinant DNA construct is useful for improving plant tolerance to cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or pests, for manipulating growth rate in plant cells by modification of the cell cycle pathway, for conferring increased resistance to plant disease, for producing galactomannan, lignin or plant growth regulators, for increasing the rate of homologous recombination in plants, for improving yield by modification of photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake or by providing improved plant growth and development under at least one stress condition or for modifying seed oil or protein yield and/or content. This is the amino acid sequence of a plant full length insert

CC polypeptide that can be used in the recombinant DNA construct of the invention.  
CC  
XX  
SQ Sequence 646 AA;  
Query Match 100.0%; Score 53; DB 8; Length 646;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
| | | | | | | | | |  
Db 4 AAAAAAAAAAAAAA 16

RESULT 4  
ID ADX78446 standard; protein; 333 AA.  
XX  
XX ADX78446;  
AC ADX78446;  
XX  
XX  
DT 23-MAR-2006 (revised)  
DT 21-APR-2005 (first entry)  
XX  
XX Plant full length insert polypeptide seqid 47812.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content.  
XX  
XX Zea mays.  
XX  
XX US2004034888-A1.  
XX  
XX 19-FEB-2004.  
XX  
XX  
XX 28-APR-2003; 2003US-00425114.  
XX  
XX 06-MAY-1999; 99US-00304517.  
XX  
XX 05-NOV-2001; 2001US-00985678.  
XX  
XX (LIUJ/) LIU J.  
XX (ZHOU/) ZHOU Y.  
XX (KOVA/) KOVALIC D K.  
XX (SCRE/) SCREEN S E.  
XX (TABA/) TABASKA J E.  
XX (CAOY/) CAO Y.  
XX  
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
XX pests, for conferring increased resistance to plant disease, or for  
XX improving yield.  
XX  
XX Claim 1; SEQ ID NO 47812; 15pp; English.

The invention describes a recombinant DNA construct comprising a polynucleotide consisting of a sequence encoding an amino acid sequence available in electronic form from the US patent office at ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide of the invention are also useful in physical arrays of molecules and as plant breeding markers. The recombinant DNA construct is useful for improving plant tolerance to cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or pests, for manipulating growth rate in plant cells by modification of the cell cycle pathway, for conferring increased resistance to plant disease, for producing galactomannan, lignin or plant growth regulators, for increasing the rate of homologous recombination in plants, for improving yield by modification of photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake or by providing improved plant growth and development under at least one stress condition or for modifying seed oil or protein yield and/or content. This is the amino acid sequence of a plant full length insert

CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This is the amino acid sequence of a plant full length insert  
CC polypeptide that can be used in the recombinant DNA construct of the  
CC invention.  
CC  
CC Revised record issued on 23-MAR-2006 : Corrected organism line  
CC  
XX Sequence 333 AA;  
SQ  
  
Query Match 94.3%; Score 50; DB 8; Length 333;  
Best Local Similarity 92.3%; Pred. No. 13;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAAAAA 13  
Db 157 SAWAAAAAAAAAAAA 169  
  
RESULT 5  
ABO58367  
ID ABO58367 standard; protein; 109 AA.  
XX  
AC ABO58367;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Human genome derived single exon protein #4601.  
XX  
KW Human; gene expression; single exon probe; microarray;  
KW alternative splicing event; genomic alteration.  
XX  
OS Homo sapiens.  
XX  
PN US2003194704-A1.  
XX  
PD 16-OCT-2003.  
XX  
PF 03-APR-2002; 2002US-00029386.  
XX  
PR 03-APR-2002; 2002US-00029386.  
XX  
PA (PENN/) PENN S G.  
PA (RANK/) RANK D R.  
PA (HANZ/) HANZEL D K.  
XX  
PI Penn SG, Rank DR, Hanzel DK;  
XX  
XX WPI; 2004-119264/12.  
DR  
XX  
XX New human genome-derived single exon nucleic acid probes useful for human  
XX gene expression analysis, for identifying or characterizing alternative  
XX splicing events, for assessing genomic alterations or as tools for  
XX surveying tissues.  
XX  
XX Claim 45; SEQ ID NO 32001; 80pp; English.  
PS  
XX  
XX The invention relates to a nucleic acid probe for measuring human gene  
XX expression, comprising any of the 27,400 fully defined nucleotide  
XX sequences in the specification, or their complements or fragments, and  
XX encoding at least 8 amino acids of any of the 688 amino acid sequences  
XX fully defined in the specification. The probe is a single exon probe that  
XX hybridises under high stringency conditions to a nucleic acid molecule  
XX expressed in human cells or tissues. Also included are a spatially-  
XX addressable set of single exon nucleic acid probes for measuring human  
XX gene expression (comprising a plurality of single exon nucleic acid  
XX probes cited above, where each of the plurality of probes is separately  
XX and addressably isolatable or amplifiable from the plurality), a single  
XX exon microarray for measuring human gene expression, a method of  
XX measuring human gene expression, a vector comprising the single exon

CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subscription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterising gross  
CC alternative splicing events, in detecting and characterising gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe protein of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docID=20030194704  
XX  
XX Sequence 109 AA;  
SQ  
  
Query Match 92.5%; Score 49; DB 8; Length 109;  
Best Local Similarity 92.3%; Pred. No. 6.3;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 AAAAAAAAAAAAAA 13  
Db 36 AAAAAAAAAAAAAA 48  
  
RESULT 6  
ABG15501  
ID ABG15501 standard; protein; 112 AA.  
XX  
AC ABG15501;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #15492.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US0008631.  
PF  
XX 31-MAR-2000; 2000US-00540217.  
XX  
PR 23-AUG-2000; 2000US-00649167.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
PI  
XX N-PSDB; AAS79688.  
DR  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
PT  
XX Claim 20; SEQ ID NO 45860; 103pp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC

CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 112 AA;

Query Match 92.5%; Score 49; DB 4; Length 112;  
 Best Local Similarity 92.3%; Pred. No. 6.5;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAMAAAAA 13  
 Db 91 AAVAAAAA 103

## RESULT 7

AA016493  
 ID AA016493 standard; protein; 252 AA.

AC AA016493;

DT 01-MAY-2003 (first entry)

DE Zorocrates fibronin 1 protein.

KW Spider silk; spider silk protein; fabric; suture; medical covering;  
 KW high-tech clothing; rope; reinforced plastic.

OS Zorocrates sp.

XX WO200299082-A2.

XX 12-DEC-2002.

PF 06-JUN-2002; 2002WO-US018256.

XX 06-JUN-2001; 2001US-0296184P.

PA (UYWY-) UNIV WYOMING.

XX Roth DA, Lewis RV;

XX WPI; 2003-140616/13.

DR N-PSDB; AAL51692.

XX Expressing spider silk protein in a higher plant, by contacting a plant  
 PT cell with silk protein encoding a gene linked to a gene that confers  
 PT resistance to selection agent, and selecting cells that survive when  
 PT incubated with the agent.

XX Claim 52; Fig 11; 114pp; English.

XX The invention comprises a method for expressing spider silk in a higher  
 CC plant (e.g. arabidopsis, tobacco, tubers, sunflower, canola, alfalfa,  
 CC soybean, maize, sorghum, wheat, cotton, small grains and rice). The  
 CC method is useful for expressing spider silk in a higher plant. The silk

CC produced is useful in the production of fabrics, sutures, medical  
 CC coverings, high-tech clothing, rope, reinforced plastics, and other  
 CC applications in which various combinations of strength and elasticity are  
 CC required. The present amino acid sequence represents a spider silk-  
 CC related protein

XX Sequence 252 AA;

Query Match 92.5%; Score 49; DB 6; Length 252;  
 Best Local Similarity 92.3%; Pred. No. 14;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAMAAAAA 13  
 Db 83 AAVAAAAA 95

## RESULT 8

AAE36894  
 ID AAE36894 standard; protein; 252 AA.

XX AC AAE36894;

XX 07-AUG-2003 (first entry)

DE Zorocrates species fibroin 1 protein.

KW Spider silk protein; flagelliform-like spider silk protein; FLAG; MaSp;  
 KW major ampullate spidroin; fabric; medical covering; clothing; parachute;  
 KW rope; reinforced plastic; suture; aerial device wing; sail; hand glider;  
 KW implantation biomaterial; resin product; fiber-reinforced plastic;  
 KW thermal injected plastic; fibroin 1.

OS Zorocrates sp.

XX WO2003020916-A2.

XX 13-MAR-2003.

PF 28-MAR-2002; 2002WO-US009663.

PR 29-AUG-2001; 2001US-0315529P.

PA (UYWY-) UNIV WYOMING.

XX Lewis RV, Hayashi CY, Gatesy JE, Motriuk D;

XX WPI; 2003-290190/28.

DR N-PSDB; AAD55739.

XX Novel spider silk protein e.g. major ampullate spidroin 2-like,  
 PT flagelliform (flag)-like spider silk proteins, useful for producing  
 PT fabric, sutures, medical coverings, high-tech clothing, rope.

XX Claim 58; Page 98; 99pp; English.

XX The invention relates to novel spider silk proteins e.g. major ampullate  
 CC spidroin 1-like (MaSp1) spider silk protein, major ampullate spidroin 2-  
 CC like (MaSp2) spider silk protein, flagelliform (flag)-like spider silk  
 CC proteins and spider silk proteins comprising atypical repetitive motifs.  
 CC Sequences of the invention are useful for producing fabrics, sutures,  
 CC medical coverings, high-tech clothing, rope and reinforced plastics. They  
 CC are used to make high-tech clothing, rope, sails, parachutes, wings on  
 CC aerial devices (e.g. hand gliders), flexible tie downs for electrical  
 CC components, sutures and as biomaterials for implantation (e.g. artificial  
 CC ligaments or aortic banding). Biomaterial applications of the spider silk  
 CC fibers involve use of natural and/or synthetic spider silk fibers in  
 CC sutures used in surgical procedures, including eye surgery, vascular  
 CC closure, bowel surgery, cosmetic surgery, reconstructive surgery (e.g.  
 CC nerve or tympanic membrane reconstruction) and central nervous system  
 CC surgery. Natural and synthetic spider silk fibers are also used in the  
 CC generation of antibiotic impregnated sutures and implant material and  
 CC matrix material for reconstruction of bone and connective tissue. Spider



CC silk proteins can be modified to alter various physical properties of  
 CC fibroin and different spider silk proteins. Synthetic spider silk fibers  
 CC may be mixed with various plastics and/or resins to prepare a fiber-  
 CC reinforced plastic and/or resin product. They are useful as structural  
 CC reinforcement material in thermal injected plastics. The present sequence  
 CC is Zorocrates species fibroin 1 protein. This sequence is used in the  
 CC exemplification of the invention

XX SQ Sequence 252 AA;

Query Match 92.5%; Score 49; DB 6; Length 252;  
 Best Local Similarity 92.3%; Pred. No. 14;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
 ||:|||||  
 DB 83 AAAAAAAAAAAAAA 95

RESULT 9  
 ABB65204  
 ID ABB65204 standard; protein; 285 AA.

XX AC ABB65204;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 22404.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX DR N-PSDB; ABL09307.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.

XX PS Disclosure; SEQ ID NO 22404; 2lpp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 285 AA;

Query Match 92.5%; Score 49; DB 4; Length 285;  
 Best Local Similarity 92.3%; Pred. No. 16;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
 ||:|||||  
 DB 189 AAAAAAAAAAAAAA 201

RESULT 10

ABG18917

ID ABG18917 standard; protein; 316 AA.

XX AC ABG18917;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #18908.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX DR N-PSDB; AAS83104.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX PS Claim 20; SEQ ID NO 49276; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABB00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences. ABB00010-ABG30377 represent novel human diagnostic  
 CC patent did not appear in the invention. Note: The sequence data for this  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 316 AA;

Query Match 92.5%; Score 49; DB 4; Length 316;  
 Best Local Similarity 92.3%; Pred. No. 17;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
 ||:|||||

Db 18 AAAAAAAAAAAAAA 30

RESULT 11

ABB69577

ID ABB69577 standard; protein; 337 AA.

AC ABB69577;

XX

XX 26-MAR-2002 (first entry)

XX

XX Drosophila melanogaster polypeptide SEQ ID NO 35523.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

KW

XX Drosophila melanogaster.

OS

XX

XX WO200171042-A2.

PN

XX

XX 27-SEP-2001.

PD

XX

XX 23-MAR-2001; 2001WO-US009231.

PF

XX

XX 23-MAR-2000; 2000US-0191637P.

PR

XX 11-JUL-2000; 2000US-00614150.

PR

XX (PEKE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

PI

XX

XX WPI; 2001-656860/75.

DR

XX N-PSDB; ABL13680.

DR

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

XX

XX Disclosure; SEQ ID NO 35523; 21pp + Sequence Listing; English.

PS

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

XX Sequence 337 AA;

SQ

Query Match 92.5%; Score 49; DB 4; Length 337;

Best Local Similarity 92.3%; Pred No. 18;

Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAAAAAA 13

Db 256 AAAAAAAAAAAAAA 268

RESULT 12

AAE35778

ID AAE35778 standard; protein; 373 AA.

XX

XX AAE35778;

AC

XX 17-JUN-2003 (first entry)

DT

XX Human SECP-30 protein.

DE

XX Human; secreted protein; SECP; cell proliferative disorder; cirrhosis;

KW

KW cancer; arteriosclerosis; hepatitis; psoriasis; developmental disorder;

KW atherosclerosis; seizure disorder; renal tubular acidosis; anaemia;

KW cataract; sensorineural hearing loss; neurological disorder; epilepsy;

KW ischaemic cerebrovascular disease; Alzheimer's disease; Pick's disease;

KW Huntington's disease; stroke; dementia; Parkinson's disease; anxiety;

KW amyotrophic lateral sclerosis; schizophrenic disorder; mental disorder;

KW Tourette's disorder; muscular dystrophy; autoimmune disorder; mood; AIDS;

KW inflammatory disorder; acquired immunodeficiency syndrome; allergy; ARDS;

KW adult respiratory distress syndrome; diabetes mellitus; Crohn's disease;

KW asthma; autoimmune thyroiditis; glomerulonephritis; rheumatoid arthritis;

KW atopic dermatitis; ulcerative colitis; trauma; cardiovascular disorder;

KW infection; congestive heart failure; heart disease; angina pectoris;

KW myocardial infarction; myocarditis; transgenic animal; gene therapy;

KW transgenic; vaccine; virucide; antibacterial; fungicide; antiparasitic;

KW protozoacide.

XX

XX Homo sapiens.

OS

XX

XX Key Location/Qualifiers

FT Peptide 1..48

FT /label= Signal\_peptide\_#1

FT Peptide 17..48

FT /label= Signal\_peptide\_#2

FT Protein 49..373

FT /note= "Mature SECP-30 protein"

XX

XX WO200297035-A2.

XX

XX 05-DEC-2002.

PD

XX

XX 21-MAY-2002; 2002WO-US016234.

PF

XX

XX 25-MAY-2001; 2001US-0293728P.

PR

XX 08-JUN-2001; 2001US-0297019P.

PR

XX 19-JUN-2001; 2001US-0299297P.

PR

XX 22-JUN-2001; 2001US-0300537P.

PR

XX 29-JUN-2001; 2001US-0301936P.

PR

XX 06-MAR-2002; 2002US-0362439P.

PR

XX 08-MAR-2002; 2002US-0363649P.

PR

XX 19-MAR-2002; 2002US-0366041P.

PR

XX (INCY-) INCYTE GENOMICS INC.

PA

XX

XX Yue H, Lee EA, Becha SD, Baughn MR, Yao MG, Tang YT;

PI Au-Young JK, Lal PG, Warren BA, Duggan BM, Tran UK, Xu Y;

PI Thangavelu K, Richardson TW, Bandman O, Jones KA, Yang J, Khan FA;

PI Emerling BM, Swarnakar A, Luo W, Wallia NK, Azimzai Y, Honchell CD;

PI Lu DAM, Griffin JA, Lee SY, Burford N, Elliott VS, Honchell CD;

PI He A, Mason PM, Li JX, Hafalia AJA, Gururajan R;

XX

XX WPI; 2003-129519/12.

DR

XX N-PSDB; AAD54316.

DR

XX

XX Novel human secreted proteins and genes encoding the proteins, useful for

PT treating, diagnosing and preventing cell proliferative,

PT autoimmune/inflammatory, cardiovascular, developmental or neurological

PT disorders.

XX

XX Claim 85; Page 168-169; 192pp; English.

PS

XX

XX The present invention relates to novel human secreted proteins (SECP) and

CC polynucleotides encoding such proteins. SECP sequences are useful for

CC diagnosing, treating and preventing cell proliferative disorders

CC including cancer (e.g. arteriosclerosis, cirrhosis, hepatitis, psoriasis

CC and atherosclerosis), developmental disorders (e.g. seizure disorders,

CC renal tubular acidosis, anaemia, cataract and sensorineural hearing

CC loss), neurological disorders (e.g. epilepsy, ischaemic cerebrovascular

CC disease, Alzheimer's disease, Pick's disease, Huntington's disease,

CC stroke, dementia, Parkinson's disease, amyotrophic lateral sclerosis,

CC schizophrenic disorders, mental disorders including mood and anxiety,

CC Tourette's disorder and muscular dystrophy), autoimmune/inflammatory

CC disorders (e.g. acquired immunodeficiency syndrome (AIDS), allergy, adult

CC respiratory distress syndrome (ARDS), asthma, autoimmune thyroiditis,

CC diabetes mellitus, Crohn's disease, glomerulonephritis, rheumatoid  
 CC arthritis, atopic dermatitis, ulcerative colitis, trauma and viral,  
 CC bacterial, fungal, parasitic, protozoal and helminthic infections) and  
 CC cardiovascular disorders (e.g. congestive heart failure, ischaemic heart  
 CC disease, angina pectoris, myocardial infarction, hypertensive heart  
 CC disease, congenital heart disease and myocarditis). They are useful for  
 CC creating knocking humanised animals or transgenic animals to model human  
 CC diseases. Sequences of the invention are useful in somatic or germline  
 CC gene therapy and in diagnostic purposes. They are also used as vaccines.  
 CC The present sequence is human SECP-30 protein

XX SQ Sequence 373 AA;

Query Match 92.5%; Score 49; DB 6; Length 373;  
 Best Local Similarity 92.3%; Pred. No. 20;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAAAAA 13  
 ||:|||||||  
 Db 36 AAAAAAAAAAAAA 48

RESULT 13

AD061835  
 ID ADO61835 standard; protein; 377 AA.

XX AC ADO61835;

XX DT 15-JUL-2004 (first entry)

XX DE Transcription factor G1850, SEQ ID 302.

XX KW Plant; transcription factor; transgenic plant; abiotic stress tolerance;  
 XX KW osmotic stress tolerance; cold tolerance; heat tolerance;  
 XX KW low nitrogen tolerance; low phosphate tolerance; fungal disease;  
 XX KW glyphosate resistance; flowering; fertility; seed development.

XX OS Arabidopsis thaliana.

XX PN WO2004031349-A2.

XX PD 15-APR-2004.

XX PF 18-SEP-2003; 2003WO-US030292.

XX PR 18-SEP-2002; 2002US-0411837P.

XX PR 17-DEC-2002; 2002US-0434166P.

XX PR 24-APR-2003; 2003US-0465809P.

XX PA (MEND-) MENDEL BIOTECHNOLOGY INC.

XX PI Jiang C, Heard JE, Ratcliffe O, Creelman RA, Adam LJ, Reuber TL;  
 PI Riechmann JL, Haake V, Dubell AN, Keddle JS, Sherman BK;

XX DR WPI; 2004-330163/30.

XX DR N-PSDB; ADO61834.

XX PT New recombinant polynucleotide encoding transcription factor  
 PT polypeptides, useful for producing transgenic plants with advantageous  
 PT properties compared to a reference plant.

XX PS Disclosure; SEQ ID NO 302; 510pp; English.

XX CC The present invention relates to novel plant transcription factor  
 CC proteins (I) and nucleotide sequences (II) (ADO61834-ADO63778). The  
 CC sequences can be used to produce transgenic plants, which overexpress  
 CC (II), where the transgenic plant has an altered trait as compared to a  
 CC non-transgenic plant or wild-type plant. The transgenic plant comprises  
 CC an altered trait selected from increased tolerance to abiotic stress,  
 CC increased tolerance to osmotic stress, increased tolerance to cold,  
 CC increased germination in cold, increased tolerance to heat, increased  
 CC germination in heat, increased tolerance to freezing conditions,  
 CC increased tolerance to low nitrogen conditions, increased tolerance to

CC low phosphate conditions, increased tolerance to disease, including  
 CC fungal disease and particularly Bryopsis, Fusarium and Botrytis,  
 CC increased tolerance to multiple fungal pathogens, increased resistance to  
 CC glyphosate, increased sensitivity to ABA, reduced sensitivity to ABA,  
 CC increased sensitivity to ACC, altered sugar sensing, increased tolerance  
 CC to sugars, altered carbon/nitrogen sensing, early flowering, late  
 CC flowering, altered flower structure, loss of flower determinacy, reduced  
 CC fertility, altered shoot meristem development, altered branching pattern,  
 CC altered stem morphology, altered vascular tissue structure, reduced  
 CC apical dominance, altered trichome density, altered trichome development,  
 CC altered trichome structure, altered root development, altered seed  
 CC avoidance, altered seed development, altered seed ripening, altered seed  
 CC germination, slow growth, fast growth, altered cell differentiation,  
 CC altered cell proliferation, altered cell expansion, altered phase change,  
 CC altered senescence, abnormal embryo development, altered programmed cell  
 CC death, lethality when overexpressed, altered necrosis patterns, increased  
 CC plant size, increased biomass, large seedlings, dwarfed plants, dark  
 CC green leaves, change in leaf shape, increased leaf size and mass, light  
 CC green or gray leaves, glossy leaves, altered abaxial/adaxial polarity,  
 CC altered seed coloration, altered seed size, altered seed shape, large  
 CC seed, increased leaf wax, increased leaf fatty acids, altered seed oil  
 CC content, altered seed protein content, altered seedprenyl content,  
 CC altered leaf prenol lipid content, increased anthocyanin levels, and  
 CC decreased anthocyanin levels. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 377 AA;

Query Match 92.5%; Score 49; DB 8; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 20;

XX Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAAAAA 13  
 |||||

Db 155 AAAAAAAAAAAAA 166

RESULT 14

ABP65077

ID ABP65077 standard; protein; 378 AA.

XX AC ABP65077;

XX DT 12-NOV-2002 (first entry)

XX DE Hypoxia-induced protein #17.

XX KW Cytostatic; vasotropic; tranquiliser; antiatherosclerotic; gene therapy;  
 KW antiinflammatory; vulnery; gynecological; ophthalmological; vaccine;  
 KW hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;  
 KW ischaemic condition; reperfusion injury; retinopathy; neonatal stress;  
 KW preclampsia; atherosclerosis; inflammatory condition; wound healing;  
 KW inflammation; erythropoiesis; hair loss; human.

XX OS Homo sapiens.

XX PN WO200246465-A2.

XX PD 13-JUN-2002.

XX PF 10-DEC-2001; 2001WO-GB005458.

XX PR 08-DEC-2000; 2000GB-00030076.

XX PR 08-FEB-2001; 2001GB-00003156.

XX PR 25-OCT-2001; 2001GB-00025666.

XX PA (OXFO-) OXFORD BIOMEDICA UK LTD.

XX PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;

XX PI Rayner WN;





CC sequence data for this patent is not represented in the printed  
 CC specification, but is based on sequence information supplied by the  
 CC European Patent Office.

SQ Sequence 512 AA;  
 Query Match 92.5%; Score 49; DB 7; Length 512;  
 Best Local Similarity 92.3%; Pred. No. 27;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 ||:|||||||  
 Db 136 AAAAAAAAAAAAAA 148

RESULT 18  
 ABB59750  
 ID ABB59750 standard; protein; 578 AA.  
 XX  
 AC ABB59750;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 6042.  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 FN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL03853.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 6042; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 578 AA;  
 Query Match 92.5%; Score 49; DB 4; Length 578;  
 Best Local Similarity 92.3%; Pred. No. 30;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 ||:|||||||  
 Db 274 AAAAAAAAAAAAAA 286

RESULT 19  
 AAB92956  
 ID AAB92956 standard; protein; 708 AA.  
 XX  
 AC AAB92956;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human protein sequence SEQ ID NO:11642.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX  
 PS Claim 8; SEQ ID NO 11642; 2537pp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesising 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 708 AA;  
 Query Match 92.5%; Score 49; DB 4; Length 708;  
 Best Local Similarity 92.3%; Pred. No. 36;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 ||:|||||||  
 Db 4 AAAAAAAAAAAAAA 16

```

XX OS Drosophila melanogaster.
XX PN WO200277023-A2.
XX PD 03-OCT-2002.
XX PF 01-FEB-2002; 2002WO-CH0000063.
XX PR 23-MAR-2001; 2001US-0277976P.
XX PA (UYZU-) UNIV ZUERICH.
XX PI Kramps T, Basler K;
XX DR WPI; 2003-018884/01.
XX DR N-PSDB; ABQ83200.
XX PT Nucleic acid sequences of the Drosophila melanogaster Daughter of Legless
XX PT gene and its encoded polypeptide, useful for developing therapeutic or
XX PT diagnostic compound for treating or diagnosing e.g. tumors or cancerous
XX PT conditions.
XX PS Claim 15; Fig 1; 68pp; English.
XX CC The present invention describes a nucleic acid sequence and its encoded
XX CC polypeptide, which are part of at least one signalling pathway in insects
XX CC and vertebrates. The nucleic acid sequence is the daughter of legless
XX CC (dollar) gene, as well as its homologues, fragments, derivatives, or
XX CC functional or structural analogues. The polypeptide is the daughter of
XX CC legless (DOLL) protein, as well as its homologues, fragments,
XX CC derivatives, or functional or structural analogues. Doll sequences have
XX CC cytotatic activity, and can be used in gene therapy. The doll nucleic
XX CC acid, DOLL protein, or their homologues, derivatives or fragments can be
XX CC used for developing a therapeutic and diagnostic compound (e.g.
XX CC antibodies or its fragments, doll antisense DNA or RNA, doll double-
XX CC stranded RNA, or chemical or naturally occurring compounds interfering
XX CC with doll function) for the treatment or diagnosis of disorders of cell
XX CC fate, differentiation or proliferation. Fragments of the doll DNA
XX CC sequences is useful as a hybridisation probe. The disorders which can be
XX CC treated using doll sequences includes human tumours, (pre-)neoplastic,
XX CC (non-)malignant or cancerous conditions. The present sequence represents
XX CC Drosophila melanogaster Doll (dollar) from the present invention
XX SQ Sequence 815 AA;

Query Match 92.5%; Score 49; DB 6; Length 815;
Best Local Similarity 92.3%; Pred. No. 42;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
Db 53 AAAAAAAAAAAAAA 65

RESULT 22
ABP71577
ID ABP71577 standard; protein; 855 AA.
XX AC ABP71577;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 41523.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.

Query Match 92.5%; Score 49; DB 4; Length 815;
Best Local Similarity 92.3%; Pred. No. 42;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13
Db 53 AAAAAAAAAAAAAA 65

RESULT 21
ABP54340
ID ABP54340 standard; protein; 815 AA.
XX AC ABP54340;
XX DT 16-JAN-2003 (first entry)
XX DE Drosophila melanogaster daughter of legless (Doll) protein SEQ ID NO:6.
XX KW Daughter of legless; Doll; wingless signalling pathway; cytostatic;
gene therapy; cancer; tumour; neoplastic.

```

XX 23-MAR-2001; 2001WO-US009231.  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL15680.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Disclosure; SEQ ID NO 41523; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX SQ Sequence 855 AA;  
 Query Match 92.5%; Score 49; DB 4; Length 855;  
 Best Local Similarity 92.3%; Pred. No. 43;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 ||:|||||||  
 Db 108 AAAAAAAAAAAAAA 120  
 RESULT 23  
 ABU64278  
 ID ABU64278 standard; protein; 892 AA.  
 AC ABU64278;  
 XX  
 XX 11-MAR-2004 (first entry)  
 XX Murine ubiquitin ligase POSH protein.  
 DE  
 DE Human; fruit fly; mouse; POSH; plenty of SH3 domains; ubiquitin ligase;  
 KW cell proliferation; enzyme; cytosolic; cancer.  
 KW  
 XX Mus sp.  
 OS  
 XX WO2003078601-A2.  
 PN  
 XX 25-SEP-2003.  
 PD  
 XX 17-MAR-2003; 2003WO-US008194.  
 PF  
 XX 15-MAR-2002; 2002US-0364530P.  
 PR 26-JUN-2002; 2002US-0391629P.  
 PR 27-NOV-2002; 2002US-0429916P.  
 XX  
 XX (PROT-) PROTEOLOGICS INC.  
 PA  
 XX Alroy I, Greener T, Ben-Avraham D;  
 PI WPI; 2003-779129/73.  
 XX N-PSDB; AAL56226.  
 DR  
 XX

PT Identifying an anti-proliferative agent, useful for treating neoplasia,  
 PT comprises providing a POSH polypeptide and a test agent, and identifying  
 PT the test agent that interacts with the POSH polypeptide.  
 XX Claim 2; Fig 15; 151pp; English.  
 XX The present invention relates to a method of identifying an anti-  
 CC proliferative agent, comprising providing a POSH (plenty of SH3 domains)  
 CC polypeptide and a test agent, and identifying the test agent that  
 CC interacts with the POSH polypeptide. The POSH protein is a ubiquitin  
 CC ligase enzyme. The method is useful for identifying agents for treating a  
 CC neoplastic condition, preferably cancer selected from thyroid carcinoma,  
 CC liver cancer (hepatocellular cancer), lung cancer, cervical cancer,  
 CC ovarian cancer, renal cell carcinoma, lymphoma, osteosarcoma, liposarcoma,  
 CC leukaemia, breast carcinoma and breast adenocarcinoma. The present  
 CC sequence is a protein sequence shown in the exemplification of the  
 CC invention  
 XX  
 XX SQ Sequence 892 AA;  
 Query Match 92.5%; Score 49; DB 7; Length 892;  
 Best Local Similarity 92.3%; Pred. No. 45;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 ||:|||||||  
 Db 419 AAAAAAAAAAAAAA 431  
 RESULT 24  
 ADH73564  
 ID ADH73564 standard; protein; 892 AA.  
 XX  
 XX ADH73564;  
 XX  
 XX 25-MAR-2004 (first entry)  
 DT  
 XX Mouse POSH protein sequence, SEQ ID 9.  
 DE  
 XX Virucide; ligase inhibitor; gene therapy; POSH; plenty of SH3 domains;  
 KW RING domain; envelop virus; retroid virus; RNA virus; retrovirus;  
 KW rhabdovirus; lentivirus; filovirus.  
 XX  
 XX Mus sp.  
 OS  
 XX EP1310552-A2.  
 PN  
 XX 14-MAY-2003.  
 PD  
 XX 11-NOV-2002; 2002EP-00257796.  
 PF  
 XX 09-NOV-2001; 2001US-0345846P.  
 PR 15-MAR-2002; 2002US-0364530P.  
 PR  
 XX (PROT-) PROTEOLOGICS INC.  
 PA  
 XX Alroy I, Tuvia S, Greener T, Ben-Avraham D;  
 PI WPI; 2003-450982/43.  
 XX N-PSDB; ADH73563.  
 DR  
 XX New POSH nucleic acids and polypeptides useful for treating viral  
 PT disorders, particularly disorders caused by envelop viruses, retroid  
 PT viruses and RNA viruses, including retroviruses, rhabdoviruses,  
 PT lentiviruses and filoviruses.  
 XX  
 XX Claim 49; Fig 15; 85pp; English.  
 PS  
 XX The invention relates to an isolated nucleic acid encoding a POSH (plenty  
 CC of SH3 domains) polypeptide. The polypeptide sequence additionally  
 CC comprises a RING domain. The nucleic acid mitigates a POSH loss of  
 CC function phenotype in a cell. The POSH loss of function phenotype is a  
 CC decrease in HIV virus like particle production. The method of the



CC invention involves producing a recombinant POSH polypeptide. Also disclosed is a method for identifying an antiviral or antiapoptotic agent, comprising providing a POSH polypeptide and a test agent, and identifying a test agent that interacts with the POSH polypeptide. The POSH polypeptide is expressed in a cell, and is a purified polypeptide. CC The POSH polypeptide comprises an SH3 or RING domain, where the test agent binds. The test agent is a polypeptide, an antibody, a small molecule or a peptidomimetic. The test agent that interacts with POSH decreases the maturation of a virus containing the PTAP or PPEY motif. CC The interaction with POSH decreases the interaction between POSH and Gag polypeptide. Alternatively, identifying an antiviral agent comprises providing a POSH nucleic acid and a test agent, and identifying a test agent that binds to the POSH nucleic acid. The test agent is a ribonucleic acid, an antisense oligonucleotide, a RNAi construct, a DNA enzyme or a ribozyme. The nucleic acids, polypeptides, compositions and methods of the invention are useful for treating viral disorders, CC particularly disorders caused by envelop viruses, retroviruses and RNA viruses, including retroviruses, rhabdoviruses, lentiviruses and CC filoviruses. The current sequence represents the mouse POSH protein CC sequence.

SQ Sequence 892 AA;

Query Match 92.5%; Score 49; DB 7; Length 892;  
Best Local Similarity 92.3%; Pred. No. 45;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||:|||||  
Db 419 AAAAAAAAAAAAAA 431

RESULT 25

ADR89697  
ID ADR89697 standard; protein; 892 AA.

XX ADR89697;

XX 18-NOV-2004 (first entry)

DT Murine POSH (Plenty of SH3 domains) coding sequence.

DE Mouse; POSH; plenty of SH3 domains; virucide; anti-HIV; cytostatic;  
KW vaccine.

KW Mus sp.

OS WO2004073609-A2.

PN 02-SEP-2004.

PD 05-FEB-2004; 2004WO-US003600.

PF 05-FEB-2003; 2003US-0445534P.

PR 03-MAR-2003; 2003US-0451437P.

PR 21-APR-2003; 2003US-0464285P.

PR 16-SEP-2003; 2003US-0503931P.

XX (PROT-) PROTEOLOGICS INC.

XX Alroy I, Taglicht DN, Reiss Y, Yaar L, Tuvia S;

XX WPI; 2004-635468/61.

XX N-PSDB; ADR89696.

XX New complex comprising a Plenty of SH3 (POSH) polypeptide and a POSH-

XX associated kinase (POSH-AK) or its subunit, useful in preparing a

XX composition for treating or preventing a POSH associated cancer.

XX Disclosure; SEQ ID NO 9; 163pp; English.

XX The present sequence is the protein sequence of murine Plenty of SH3

XX domains (POSH). The invention provides novel complexes of POSH

CC polypeptides and POSH associated kinases (POSH-AKs). An isolated, purified or recombinant complex comprising a POSH polypeptide and a POSH-AK or a subunit of a POSH-AK is claimed. The complex is used in methods for identifying agents that modulate an activity of POSH or POSH-AK by disrupting a complex of POSH and POSH-AK. These modulator agents can be used as antiviral agents, especially where the virus is an envelope virus, HIV or West Nile virus, or as anti-apoptotic or anti-cancer agents in which a modulator agent will decrease the proliferation or survival of a POSH-associated cancer cell. The modulator agents may also inhibit trafficking of a protein through the secretory pathway. An inhibitor of POSH, e.g. an agent that inhibits a POSH activity such as ubiquitin ligase activity or interaction with a POSH-AP, or an agent that inhibits POSH expression, can be used to inhibit the Jun kinase (JNK) pathway in a human cell and hence to treat a JNK pathway-associated disease. A POSH inhibitor may also be used to treat a protein kinase A (PKA) associated disease. A method of enhancing interaction of a POSH polypeptide with a second protein in a cell comprises contacting the cell with an agent that inhibits phosphorylation of POSH by PKA. An isolated, purified or recombinant phosphorylated POSH polypeptide is claimed, which is phosphorylated at a consensus PKA phosphorylation site, at a site of sequence K/R-R-X-S/T-hydrophobic or R-X-X-S/T-hydrophobic. Also claimed is a portion of a POSH polypeptide comprising 15-100 consecutive amino acids of a mammalian POSH polypeptide including a consensus PKA phosphorylation site and comprising at least one modified amino acid or peptidomimetic moiety, which inhibits PKA phosphorylation of POSH.

XX Sequence 892 AA;

Query Match 92.5%; Score 49; DB 8; Length 892;  
Best Local Similarity 92.3%; Pred. No. 45;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13

||:|||||  
Db 419 AAAAAAAAAAAAAA 431

RESULT 26

ADS34245

ID ADS34245 standard; protein; 892 AA.

XX ADS34245;

XX 02-DEC-2004 (first entry)

XX POSH protein associated protein #4.

XX cytostatic; nontropic; neuroprotective; antiparkinsonian; anticonvulsant;

KW antiviral; neuroleptic; central nervous system; POSH polypeptide;

KW POSH-associated protein; POSH-AP; HERPUDI; Ubiquitin ligase;

KW antiviral agent; anti-apoptotic agent; anti-cancer agent;

KW secretory pathway trafficking inhibitor;

KW neurological disorder progression disorder; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease; schizophrenia;

KW Niemann-Pick's disease.

OS Homo sapiens.

XX WO2004078130-A2.

XX 16-SEP-2004.

XX 02-MAR-2004; 2004WO-US006308.

XX 03-MAR-2003; 2003US-0451437P.

XX 05-MAR-2003; 2003US-0452284P.

XX 19-MAR-2003; 2003US-0455760P.

XX 20-MAR-2003; 2003US-0456640P.

XX 03-APR-2003; 2003US-0460526P.

XX 04-APR-2003; 2003US-0460792P.

XX 21-APR-2003; 2003US-0464285P.

XX 09-MAY-2003; 2003US-0469462P.

XX 15-MAY-2003; 2003US-0471378P.

PR 20-MAY-2003; 2003US-0472327P.  
 PR 30-MAY-2003; 2003US-0474706P.  
 PR 03-JUN-2003; 2003US-0475825P.  
 PR 17-JUN-2003; 2003US-0479317P.  
 PR 19-JUN-2003; 2003US-0480215P.  
 PR 19-JUN-2003; 2003US-0480376P.  
 PR 08-AUG-2003; 2003US-0493860P.  
 PR 28-AUG-2003; 2003US-0498634P.  
 PR 16-SEP-2003; 2003US-0503931P.  
 PR 10-NOV-2003; 2003WO-US035712.  
 PR 05-FEB-2004; 2004WO-US003600.  
 PR 02-MAR-2004; 2004US-0549896P.  
 PA (PROT-) PROTEOLOGICS INC.  
 XX  
 XX Taglicht DN, Alroy I, Reiss Y, Yaar L, Ben-Avraham D, Tuvia S;  
 PI Greener T;  
 XX WPI; 2004-662346/64.  
 XX  
 XX Isolated, purified or recombinant complex, useful for identifying an  
 PT antiviral, anti-apoptotic or anti-cancer, comprising POSH polypeptide and  
 PT POSH-associated protein (POSH-AP).  
 XX  
 XX Disclosure; SEQ ID NO 9; 374pp; English.  
 XX  
 XX The invention relates to an isolated, purified or recombinant complex (I)  
 CC comprising a POSH polypeptide and a POSH-associated protein (POSH-AP) (a)  
 CC or HERPUD1 and a Ubiquitin ligase (b). Methods using (I), (a) or (b) are  
 CC useful for identifying an agent that modulates an activity of a POSH  
 CC polypeptide or POSH-AP, for identifying an antiviral agent, an anti-  
 CC apoptotic agent, an anti-cancer agent, an agent that inhibits trafficking  
 CC of a protein through the secretory pathway, an agent that inhibits the  
 CC progression of a neurological disorder, an agent that modulates a POSH  
 CC function, an agent that modulates a HERPUD1 function. The methods can be  
 CC used for treating a viral infection, for inhibiting an activity of a POSH  
 CC -AP in a cell, for treating a POSH-associated disease in a subject. The  
 CC POSH-associated disease is viral infection, POSH-associated cancer or  
 CC POSH-associated neurological disorder. The methods are useful for  
 CC treating or preventing POSH-associated neurological disorder in a subject  
 CC e.g. Alzheimer's disease, Parkinson's disease, Huntington's disease,  
 CC schizophrenia, Niemann-Pick's disease. This sequence corresponds to a  
 CC protein of the invention.  
 XX  
 XX Sequence 892 AA;  
 SQ  
 Query Match 92.5%; Score 49; DB 8; Length 892;  
 Best Local Similarity 92.3%; Pred. No. 45;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 Db 419 AAAAAAAAAAAAAA 431  
 RESULT 27  
 ADT77904  
 ID ADT77904 standard; protein; 892 AA.  
 XX  
 XX ADT77904;  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 XX Mouse POSH (plenty of SH3 domains) polypeptide.  
 DE  
 XX  
 XX Mouse; POSH; plenty of SH3 domains; ubiquitin ligase; enzyme;  
 KW neuroprotective; nootropic; antiparkinsonian; anticonvulsant;  
 KW cerebroprotective; CNS-gen.; antidepressant; neuroleptic; gene therapy.  
 XX  
 OS Mus sp.  
 XX  
 XX WO2004089302-A2.  
 PN  
 XX

PD 21-OCT-2004.  
 XX  
 PF 05-APR-2004; 2004WO-US010582.  
 XX  
 PR 03-APR-2003; 2003US-0460526P.  
 PR 03-JUN-2003; 2003US-0475825P.  
 PR 02-MAR-2004; 2004WO-US006308.  
 XX  
 PA (PROT-) PROTEOLOGICS INC.  
 XX  
 XX Yaar L, Alroy I, Reiss Y, Taglicht DN;  
 XX WPI; 2004-748573/73.  
 DR N-PSDB; ADT77903.  
 XX  
 XX New complex of POSH polypeptides and POSH-associated protein (POSH-AP)  
 PT useful for identifying an agent for treating neurological disorder e.g.  
 PT Alzheimer's disease, Parkinson's disease.  
 XX  
 XX Disclosure; SEQ ID NO 9; 145pp; English.  
 XX  
 XX The present sequence is the protein sequence of murine POSH (plenty of  
 CC SH3 domains), a ubiquitin ligase. The invention relates to POSH and to  
 CC the discovery of novel interactions between POSH and POSH associated  
 CC proteins, especially HERPUD1. Methods are provided for modulating a  
 CC process in which POSH participates by targeting HERPUD1 or the  
 CC POSH:HERPUD1 interaction, or for modulating a process in which HERPUD1  
 CC participates by targeting POSH. A claimed method of treating a  
 CC neurological disorder comprises administering an agent that inhibits a  
 CC ubiquitin ligase activity of POSH and hence the POSH-mediated  
 CC ubiquitination of HERPUD1, without substantially inhibiting POSH auto-  
 CC ubiquitination. The neurological disorder is selected from: Alzheimer's  
 CC disease, Parkinson's disease, Huntington's disease, Pick's disease,  
 CC Niemann-Pick's disease, a prion-associated disease, depression and  
 CC schizophrenia. The modulator agent is a short interfering RNA (siRNA)  
 CC construct, a small molecule, an antibody or an antisense construct.  
 XX  
 XX Sequence 892 AA;  
 SQ  
 Query Match 92.5%; Score 49; DB 8; Length 892;  
 Best Local Similarity 92.3%; Pred. No. 45;  
 Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AAAAAAAAAAAAAA 13  
 Db 419 AAAAAAAAAAAAAA 431  
 RESULT 28  
 ADU73761  
 ID ADU73761 standard; protein; 892 AA.  
 XX  
 XX ADU73761;  
 XX  
 DT 10-FEB-2005 (first entry)  
 XX  
 XX Mouse POSH polypeptide.  
 DE  
 XX  
 XX POSH; Cbl-b-associated protein; Virucide; Anti-HIV; Cytostatic;  
 KW Immunosuppressive; Neuroprotective; Nootropic; HIV-1 infection;  
 KW HIV-2 infection; cancer; neurological disease; gene therapy.  
 XX  
 OS Mus sp.  
 XX  
 XX WO2004099388-A2.  
 PN  
 XX  
 PD 18-NOV-2004.  
 XX  
 PF 05-MAR-2004; 2004WO-US006619.  
 XX  
 XX 05-MAR-2003; 2003US-0452284P.  
 PR 20-MAR-2003; 2003US-0456640P.  
 PR 09-MAY-2003; 2003US-0469462P.  
 PR

PR 15-MAY-2003; 2003US-0471378P.  
PR 19-JUN-2003; 2003US-0480215P.  
PR 19-JUN-2003; 2003US-0480376P.  
XX (PROT-) PROTEOLOGICS INC.  
XX  
XX Reiss Y, Taglicht DN, Alroy I, Tuvia S, Barr HM;  
PI  
XX WPI; 2004-805123/79.  
XX N-PSDB; ADU73760.  
XX  
XX New isolated, purified or recombinant complex comprises a Cbl-b  
PT polypeptide, a POSH polypeptide, useful for treating Cbl-associated  
PT diseases, e.g. cancer, viral infection, disorders of the immune system,  
PT or neurological disorders.  
XX  
XX Disclosure; SEQ ID NO 9; 190pp; English.  
XX  
CC The present sequence is the protein sequence of mouse POSH. The invention  
CC relates to novel associations between Cbl-b and POSH. Cbl-b and POSH are  
CC involved in viral maturation, including the production, post-  
CC translational processing, assembly and/or release of proteins in a viral  
CC particle. Viral infections may be ameliorated by inhibiting an activity  
CC (e.g. ubiquitin ligase activity or target protein interaction) of POSH or  
CC Cbl-b. The virus is a retrovirus, an RNA virus or an envelope virus,  
CC especially a human immunodeficiency virus. A claimed method of  
CC identifying an antiviral agent comprises identifying a test agent that  
CC disrupts a complex of a Cbl-b polypeptide and a POSH polypeptide. A  
CC claimed method of treating a viral infection comprises administering an  
CC agent that inhibits the expression or activity of Cbl-b. The agent is  
CC preferably an siRNA construct ADU73811-ADU73816 that inhibits expression  
CC of human Cbl-b, an antisense construct, an antibody, a polypeptide or a  
CC small molecule. Also claimed are: methods of identifying and evaluating  
CC agents that modulate a Cbl-b or POSH function by measuring the effect of  
CC the agent on Cbl-b-mediated or POSH-mediated ubiquitination or on  
CC budding, release, infectivity or reverse transcriptase activity of a  
CC virus or virus-like particle; methods of identifying an anti-apoptotic  
CC agent, an anti-cancer agent or an agent that inhibits progression of a  
CC neurological disorder, where the agent disrupts a complex comprising Cbl-  
CC b and POSH (or a Cbl-b-associated protein such as POSH); and methods of  
CC treating or preventing a POSH-associated cancer, POSH-associated  
CC neurological disorder or POSH-associated viral disorder using an agent  
CC that inhibits the expression or an activity of a Cbl-b polypeptide.  
XX  
XX Sequence 892 AA;  
XX  
XX Query Match 92.5%; Score 49; DB 8; Length 892;  
XX Best Local Similarity 92.3%; Pred. No. 45;  
XX Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 AAAAAAAAAAAAAA 13  
Db 419 AAAAAAAAAAAAAA 431  
RESULT 29  
ADU68997  
ID ADU68997 standard; protein; 892 AA.  
XX  
XX AC ADU68997;  
XX  
XX DT 10-FEB-2005 (first entry)  
XX  
XX DE House mouse plenty of SH3 domains (POSH) protein.  
XX  
XX protein secretion; protein localization; golgi apparatus;  
KW plenty of SH3 domains protein; POSH; neuroprotective; nootropic;  
KW antidiabetic; antiinflammatory; viricide; antibacterial; antiparasitic;  
KW immunosuppressive; antiparkinsonian; anticonvulsant; neuroleptic;  
KW cytosstatic; CNS-Gen.; cardiovascular disease; viral infection;  
KW bacterial infection; bacterial infection; microbial protein;  
KW neurological disease; immune disorder; Alzheimers disease;  
KW parkinsons disease; Huntingtons chorea; scleroderma;

KW niemann pick disease; prion infection; inflammation; autoimmune disease;  
KW diabetes mellitus; glycosylation; liver tumor; hepatocellular carcinoma;  
KW lung tumor; cervical dysplasia; colorectal tumor; ovary tumor;  
KW renal cell carcinoma; lymphoma; osteosarcoma; prostate tumor; leukemia;  
KW breast tumor; thyroid tumor; cancer.  
XX  
XX Mus musculus.  
XX  
XX WO2004098492-A2.  
XX  
XX 18-NOV-2004.  
XX  
XX 10-NOV-2003; 2003WO-US0357112.  
XX  
XX 11-NOV-2002; 2002EP-00257796.  
XX 12-NOV-2002; 2002US-00293965.  
PR 29-JAN-2003; 2003US-0443495P.  
PR 02-JUN-2003; 2003US-0475387P.  
PR 01-AUG-2003; 2003US-0491891P.  
XX (PROT-) PROTEOLOGICS INC.  
XX  
XX Alroy I, Reiss Y, Taglicht DN, Tuvia S, Yaar L;  
XX  
XX WPI; 2004-833580/82.  
XX N-PSDB; ADU68996.  
XX  
XX Modulating activity or localization of trans-Golgi network (TGN)-  
XX associated protein useful, eg, for inhibiting beta amyloid production or  
XX reducing virus infectivity comprises modulating activity of plenty of SH3  
XX domains protein (POSH).  
XX  
XX Example 22; SEQ ID NO 9; 248pp; English.  
XX  
XX The invention relates to a novel method for modulating the activity or  
XX localization of a trans-Golgi network (TGN)-associated protein. The  
XX method involves modulating the activity of one or more TGN-associated  
XX proteins chosen from plenty of SH3 domains protein (POSH); a POSH-pathway  
XX polypeptide, a POSH-associated polypeptide or a POSH binding protein. The  
XX method of the invention demonstrates neuroprotective, nootropic,  
XX antidiabetic, antiinflammatory, viricide, antibacterial, antiparasitic,  
XX immunosuppressive, antiparkinsonian, anticonvulsant, neuroleptic,  
XX cytostatic and CNS-Gen applications. The method of the invention may be  
XX useful for modulating the activity or localization of a TGN-associated  
XX protein. In particular, the method may be useful in inhibiting processing  
XX of a viral, bacterial, parasitic or microbial protein, and for inhibiting  
XX aberrant processing of a protein associated with a neurological disorder,  
XX cardiovascular disease or immunological disorder, in the subject, where  
XX the subject exhibits a neurological disorder chosen from Alzheimer's  
XX disease, Parkinson's disease, Huntington's disease, schizophrenia,  
XX Niemann-pick's disease and prion-associated disease, preferably an  
XX Alzheimer's disease. Alternatively, the subject may demonstrate an  
XX immunological disorder chosen from an inflammatory disease or an  
XX autoimmune disease, such as diabetes or a disease associated with  
XX improper glycosylation, where the improper glycosylation is N-linked  
XX glycosylation. Furthermore, the method may be useful for inhibiting  
XX secretion of a protein such as a myristylated protein, and for inhibiting  
XX the processing of a protein that is associated with a neoplastic  
XX condition, in a subject having cancer. The cancer is chosen from thyroid  
XX carcinoma, liver cancer, hepatocellular cancer, lung cancer, cervical  
XX cancer, colorectal cancer, ovarian cancer, renal cell carcinoma,  
XX lymphoma, osteosarcoma, prostate cancer, liposarcoma, leukemia, breast  
XX carcinoma and breast adeno-carcinoma, preferably breast cancer or  
XX colorectal cancer. The current sequence is that of a POSH protein of the  
XX invention.  
XX  
XX Sequence 892 AA;  
XX  
XX Query Match 92.5%; Score 49; DB 8; Length 892;  
XX Best Local Similarity 92.3%; Pred. No. 45;  
XX Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 AAAAAAAAAAAAAA 13

DB 419 AAAAAAAAAAAAA 431

RESULT 30  
ADW87412

ID ADW87412 standard; protein; 892 AA.

XX ADW87412;

AC ADW87412;

DT 07-APR-2005 (first entry)

XX Mouse POSH protein SEQ ID NO 9.

DE POSH inhibitor; ubiquitin ligase inhibitor; APP; cell proliferation;  
KW amyloid precursor protein; viral infection; neurological disease;  
KW Huntingtons chorea; Alzheimers disease; parkinsons disease;  
KW pick's disease; major depressive disorder; schizophrenia; cancer;  
KW cystostatic; virucide; anti-HIV; neuroprotective; nootropic;  
KW anticonvulsant; antiparkinsonian; cerebroprotective; antidepressant;  
KW neuroleptic.

OS Mus sp.

XX WO2005007141-A2.

PN 27-JAN-2005.

XX 09-JUL-2004; 2004WO-US021900.

XX 11-JUL-2003; 2003US-0486730P.

PR 24-JUL-2003; 2003US-0489795P.

PR 10-NOV-2003; 2003WO-US035712.

PR 02-MAR-2004; 2004US-0549896P.

PR 02-MAR-2004; 2004WO-US006308.

PR 05-MAR-2004; 2004WO-US006619.

PR 05-APR-2004; 2004WO-US010582.

PR 28-MAY-2004; 2004WO-US016855.

PA (PROT-) PROTEOLOGICS INC.

PI Alroy I, Tuvia S, Reiss Y, Levi-Hacham O;

XX WPI; 2005-132260/14.

DR N-PSDB; ADW87411.

XX New small molecule inhibitor of a POSH polypeptide useful e.g. to inhibit  
PT ubiquitin ligase activity of polypeptide, to treat viral infections,  
PT neurological disorder e.g. Alzheimer's disease, and to treat or prevent  
PT cancers.

XX Claim 45; SEQ ID NO 9; 171pp; English.

XX This invention describes a novel small molecule inhibitor of a POSH  
CC polypeptide chosen from human Cbl-b polypeptide, a human PEM-3-like  
CC polypeptide, a STAH1 polypeptide or a TTC3 polypeptide, comprising a RING  
CC domain. The POSH polypeptide is ARIH2, ASF1B, ATP5A1, BANP1, BAT3, BCAR1,  
CC BCU9, BIA2, BRD4, Cbl or f17, C6 or f11, C6 or f60, CBX4, CDT1, CGI-27,  
CC CIC, CL25084, CLK2, COL1A1, DAP, DDX31, DKF2p43B1231, DKF2p761A052,  
CC DLG5, DNM2, DRP2, EBF1A1, EGLN2, EIF4EBP1, EVPL, EWSR1, FAT, FL53657,  
CC FLJ10120, FLJ13231, FLJ13479, FU37147, FSTL1, GC20, GLUL, HEBP2.  
CC H8 31535, H8 380933, HSPA1B, HSPC016, HSPC142, ITGB, J03930, KHDRBS1,  
CC KIAA0391, KIAA1111, KIAA1598, LAMA5, LOC118987, LOC90987, MADH6, MAP1A,  
CC MBD2, MRPL36, MT2A, NAPA, NQO2, NUMA1, OPTN, PA1-RBP1, PAWR, PCBP1,  
CC PCNT2, PGD, PIAS3, POLQ, POLR2J2, PRDX5, PROL4, RAP80, RBAF600, RNH, RPL,  
CC RP220, RP227A, SBTB 1, SF3A2, SH2D2A, SIAH2, SLC2A1, SRPK2, SSR4, STC2,  
CC THOC2, TLE1, TPX2, UBB, UBC, VCL, XM 2 08944, XTP3TPB, ZFW1, ZNF147,  
CC ZNF151, PRA, SNX1, SNX3, ATP6V0C, PTPN12, PPP1CA, GOSR2, CENTB1, DREF1,  
CC ARF1, ARF5, PACS-1, EP86L2, HERPUD1, UNC84B, MSTP028, GOCAP, EIF3S3,  
CC SRA1, CBL-B, RALA, SIAH1, SMN1, SMN2, SYNE1, TTC3, VCV21P1, UBE2N  
CC (UBC13), ARHV (Chp), WASF1, HRP55, SPG20, HLA-A or HLA-B. The molecules  
CC of the invention are useful in the inhibition of ubiquitin ligase  
CC activity of a polypeptide or maturation of a virus in a cell, for

CC transport of amyloid precursor protein (APP) in a cell or cell  
CC proliferation or amyloid polypeptide production in a cell and also for  
CC the treatment of a viral infection (an envelope, a retroviral or a human  
CC immunodeficiency viral infections) or neurological disorder (Alzheimer's  
CC disease, Parkinson's disease, Huntington's disease, Pick's disease,  
CC depression or schizophrenia). The inhibitor is also useful in the  
CC treatment and prevention of cancer, in a method of inhibiting ubiquitin  
CC ligase activity of a polypeptide, for maturation of a virus in a cell,  
CC for transport of APP and is capable of targeting one or more proteins.  
CC The products of the invention have cytostatic, virucide, anti-HIV,  
CC neuroprotective, nootropic, anticonvulsant, antiparkinsonian,  
CC cerebroprotective, antidepressant and neuroleptic activity. This sequence  
CC represents a mouse POSH protein.

XX SQ Sequence 892 AA;

Query Match 92.5%; Score 49; DB 9; Length 892;  
Best Local Similarity 92.3%; Pred. No. 45;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAAAA 13  
||:|||||||  
DB 419 AAAAAAAAAAAAA 431

RESULT 31  
ADZ66429

ID ADZ66429 standard; protein; 892 AA.

XX ADZ66429;

XX 14-JUL-2005 (first entry)

DE Mouse POSH.

XX Virucide; human; ubiquitin ligase; Ub; plenty of SH3 domains; POSH;  
KW POSH-associated protein; POSH-AP; SH3 domain; viral maturation;  
KW ubiquitinylation; vesicular trafficking; viral infection; enzyme.

XX Mus sp.

XX WO2005038007-A2.

PN 28-APR-2005.

XX 18-OCT-2004; 2004WO-US034498.

PF 17-OCT-2003; 2003US-0512046P.

PR (PROT-) PROTEOLOGICS INC.

XX Reiss Y;

XX WPI; 2005-322863/33.

DR N-PSDB; ADZ66428.

XX Isolated, purified or recombinant complex useful for identifying anti-  
CC viral agent and anti-apoptotic agent, comprises plenty of SH3 domains  
CC POSH polypeptide and POSH-associated proteins POSH-AP.

XX Disclosure; SEQ ID NO 9; 171pp; English.

XX This sequence represents a mouse ubiquitin (Ub) ligase, plenty of SH3  
CC domains (POSH). The composition of the invention comprises a human POSH  
CC polypeptide and a POSH-associated protein (POSH-AP), or a polypeptide  
CC comprising a domain that is 90% identical to a POSH SH3 domain and a POSH  
CC -AP comprising an E2. Interactions between POSH and POSH-AP may be  
CC involved in many cellular processes, including viral maturation,  
CC stimulating ubiquitinylation of certain proteins, vesicular trafficking,  
CC regulating Rac or JNK signaling pathway, etc. An antibody which is  
CC immunoreactive with the epitope of human POSH amino acid sequence, where  
CC the antibody disrupts the interaction between a POSH polypeptide and a  
CC POSH-AP may be used for inhibiting viral infection. The POSH-AP comprises



```
Db 419 AAAAAAAAAA 431

RESULT 34
ABG05909
ID ABG05909 standard; protein; 992 AA.
XX
AC ABG05909;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #5900.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
  food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR N-PSDB; AAS70096.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
  diagnostics, forensics, gene mapping, identification of mutations
  responsible for genetic disorders or other traits and to assess
  biodiversity.
XX
PS Claim 20; SEQ ID NO 36268; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
  sequences. (I) is useful as hybridisation probes, polymerase chain
  reaction (PCR) primers, oligomers, and for chromosome and gene mapping.
  CC and in recombinant production of (II). The polynucleotides are also used
  CC in diagnostics as expressed sequence tags for identifying expressed
  CC genes. (I) is useful in gene therapy techniques to restore normal
  CC activity of (II) or to treat disease states involving (II). (II) is
  CC useful for generating antibodies against it, detecting or quantitating a
  CC polypeptide in tissue, as molecular weight markers and as a food
  CC supplement. (II) and its binding partners are useful in medical imaging
  CC of sites expressing (II). (I) and (II) are useful for treating disorders
  CC involving aberrant protein expression or biological activity. The
  CC polypeptide and polynucleotide sequences have applications in
  CC diagnostics, forensics, gene mapping, identification of mutations
  CC responsible for genetic disorders or other traits to assess biodiversity
  CC and to produce other types of data and products dependent on DNA and
  CC amino acid sequences. ABG00010-ABG030377 represent novel human diagnostic
  CC amino acid sequences of the invention. Note: The sequence data for this
  CC patent did not appear in the printed specification, but was obtained in
  CC electronic format directly from WIPO at
  CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 992 AA;

Query Match 92.5%; Score 49; DB 4; Length 992;
Best Local Similarity 92.3%; Pred. No. 50;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 13
Db 4 AAAAAAAAAA 16

RESULT 35
ABB66003
ID ABB66003 standard; protein; 1184 AA.
XX
AC ABB66003;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 24801.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
  pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL10106.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
  genes from Drosophila and for elucidating cell signalling and cell-cell
  interactions.
XX
PS Disclosure; SEQ ID NO 24801; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
  capable of detecting 1000 or more genes from Drosophila. The invention is
  useful in developmental biology and in elucidating cell signalling and
  cell-cell interactions in higher eukaryotes for the development of
  CC insecticides, therapeutics and pharmaceutical drugs. The invention
  CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
  CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
  CC ABB72072). The sequence data for this patent did not form part of the
  CC printed specification, but was obtained in electronic format directly
  CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1184 AA;

Query Match 92.5%; Score 49; DB 4; Length 1184;
Best Local Similarity 92.3%; Pred. No. 59;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 13
Db 701 AAAAAAAAAA 713

RESULT 36
ABB58493
ID ABB58493 standard; protein; 1235 AA.
XX
AC ABB58493;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2271.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
  pharmaceutical.
XX
```



DR N-PSDB; ABL03088.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 3747; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
CC ABBS72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1937 AA;  
Query Match 92.5%; Score 49; DB 4; Length 1937;  
Best Local Similarity 92.3%; Pred. No. 93;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAAAAAA 13  
Db 330 AAAAAAAAAAAAAA 342  
RESULT 39  
AAE36107  
ID AAE36107 standard; protein; 1944 AA.  
XX  
AC AAE36107;  
XX  
DT 26-JUN-2003 (first entry)  
XX  
DE Human chromodomain helicase DNA binding protein (CHD) #5.  
XX  
KW Human; p53 pathway; therapeutic; angiogenic disorder; apoptotic disorder;  
KW chromodomain helicase DNA binding protein; CHD; cancer; gene therapy;  
KW cell proliferative disorder; chromatin organisation modifier domain;  
KW cytosstatic.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 628..670  
FT /note= "Chromodomain"  
FT Domain 739..1035  
FT /note= "SNF2N domain"  
FT Domain 1100..1174  
FT /note= "Conserved C-terminal domain"  
XX  
PN WO200298899-A2.  
XX  
FD 12-DEC-2002.  
XX  
PF 03-JUN-2002; 2002WO-US017466.  
XX  
PF 05-JUN-2001; 2001US-0296076P.  
PR 10-OCT-2001; 2001US-0328603P.  
PR 22-OCT-2001; 2001US-0338733P.  
PR 15-FEB-2002; 2002US-0357253P.  
PR 15-FEB-2002; 2002US-0357600P.  
XX  
PA (EXEL-) EXELIXIS INC.  
XX  
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
PI Lioubin MN;  
XX  
XX WPI; 2003-156840/15.

XX  
PT Identifying a candidate p53 pathway-modulating agent as therapeutic  
PT targets for disorders related to defective p53 function e.g. cancer by  
PT contacting an assay system having purified CHD polypeptide or nucleic  
PT acid, with a test agent.  
XX  
PS Claim 13; Page 207-215; 278pp; English.  
XX  
CC The present invention relates to a method for identifying candidate p53  
CC pathway modulating agents. The method involves contacting an assay system  
CC comprising purified chromatin organisation modifier (chromo) domain  
CC helicase DNA binding proteins (CHD), nucleic acids, their functionally  
CC active fragments or derivatives, with a test agent under conditions  
CC where, but for the presence of the test agent, the system provides a  
CC reference activity. The methods are useful for identifying modulators of  
CC the p53 pathway as therapeutic targets for disorders associated with  
CC defective p53 function, such as angiogenic disorders, apoptotic disorders  
CC or cell proliferative disorders, e.g. cancer. The modulators are useful  
CC as research reagents, diagnostics and therapeutics. The invention is also  
CC useful in gene therapy. The present sequence is human chromodomain  
CC helicase DNA binding protein  
XX  
SQ Sequence 1944 AA;  
Query Match 92.5%; Score 49; DB 6; Length 1944;  
Best Local Similarity 92.3%; Pred. No. 94;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAAAAAA 13  
Db 207 AAAAAAAAAAAAAA 219  
RESULT 40  
ADI26773  
ID ADI26773 standard; protein; 1944 AA.  
XX  
AC ADI26773;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Human CHD3, SEQ ID 30.  
XX  
KW Cytostatic; ophthalmological; Antiarthritic; Antirheumatic;  
KW Antiinflammatory; Antiarteriosclerotic; Antipsoriatic; Gynaecological;  
KW Vasotropic; Vulnery; Hypotensive; Antianginal; Cardiant; Osteopathic;  
KW gene therapy; angiogenic; tumour vascularisation; retinopathy;  
KW rheumatoid arthritis; Crohn disease; atherosclerosis;  
KW ovarian hyperstimulation; psoriasis; endometriosis; neovascularization;  
KW restenosis; cardiovascular disease; hypertension; Raynaud disease;  
KW muscular degeneration; osteoporosis; human; CHD3.  
XX  
OS Homo sapiens.  
XX  
PN FR2837391-A1.  
XX  
PD 26-SEP-2003.  
XX  
PF 22-MAR-2002; 2002FR-00003655.  
XX  
PR 22-MAR-2002; 2002FR-00003655.  
XX  
PA (GENE-) GENE SIGNAL.  
PA (ALMA/) AL MAHMOOD S.  
XX  
PI Colin S, Schneider C, Al Mahmood S;  
XX  
XX WPI; 2003-806304/76.  
DR N-PSDB; ADI26769.  
XX  
XX Antiangiogenic composition, useful for treating e.g. tumors, comprises  
PT new angiogenesis-related nucleic acids or related polypeptides,  
PT antibodies or antisense sequences.



XX Claim 5; SEQ ID NO 34; 110pp; French.  
PS  
XX The present invention relates to antiangiogenic pharmaceutical  
XX compositions. The compositions comprise, as active ingredient, one or  
CC more of: nucleic acid (I; ADI26740-ADI26744 and ADI26766-ADI26769) from  
CC an endothelial cell gene the expression of which is induced by an  
CC angiogenic agent, or its complement or fragment; polypeptide (II;  
CC ADI26745-ADI26749 and ADI26770-ADI26773) encoded by (I), or its fragment;  
CC an antisense nucleic acid (III; ADI26750-ADI26753) able to inhibit  
CC expression of (I); or an antibody (Ab) that binds (II). The compositions  
CC are used for treating a wide range of angiogenic diseases: tumour  
CC vascularisation; retinopathy; rheumatoid arthritis; Crohn disease;  
CC atherosclerosis; ovarian hyperstimulation; psoriasis; endometriosis  
CC associated with neovascularization; restenosis (arterial or after balloon  
CC angioplasty); overgrowth of cells in wound healing; peripheral vascular  
CC disease; hypertension; vascular inflammation; Raynaud disease; aneurysm;  
CC thrombophlebitis; lymphedema; ischaemia; angina; myocardial infarct;  
CC chronic heart disease; (congestive) cardiac insufficiency; age-related  
CC muscular degeneration and osteoporosis. This sequence is chromodomain  
CC helixase DNA binding protein 3 (CHD3).  
XX  
SQ Sequence 1944 AA;  
Query Match 92.5%; Score 49; DB 7; Length 1944;  
Best Local Similarity 92.3%; Pred. No. 94;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAAA 13  
Db 207 AAAAAAAAAAAAA 219

Search completed: September 9, 2006, 22:48:07  
Job time : 101.418 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:48:41 ; Search time 17.0886 Seconds  
(without alignments)  
84.457 Million cell updates/sec

Title: US-10-617-568-1

Perfect score: 73

Sequence: 1 PVSKVRMATPLMQA 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR\_80:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73	100.0	216	1 HLHUG	class II histocomp
2	61	83.6	279	1 HLHMSG	class II histocomp
3	60	82.2	204	2 JC4796	intracellular memb
4	59	80.8	280	2 S04362	class II histocomp
5	42	57.5	327	2 G84420	TATC-like protein
6	42	57.5	340	2 T44583	conserved hypotet
7	40	54.8	92	2 T03471	probable poly(3-hy
8	40	54.8	543	2 T00513	cytochrome P450 ho
9	40	54.8	720	2 AG0918	DNA helicase II (i
10	39	53.4	554	2 T43466	hypothetical prote
11	39	53.4	650	2 JC7093	Fas associated fac
12	39	53.4	720	1 HJEC22	DNA helicase II (E
13	39	53.4	720	2 G91221	DNA-dependent ATPa
14	39	53.4	720	2 B86068	DNA-dependent ATPa
15	39	53.4	3390	1 GNWV33	genome polypeptin
16	38	52.1	168	2 T47054	hypothetical prote
17	38	52.1	184	2 AD0232	hypothetical prote
18	38	52.1	184	2 T17446	hypothetical 21.1k
19	38	52.1	628	2 B91146	probable integral
20	38	52.1	628	2 F85991	probable integral
21	38	52.1	828	2 T22367	hypothetical prote
22	38	52.1	1034	2 D65119	acriflavin resista
23	38	52.1	1046	2 S36330	multidrug-efflux t
24	38	52.1	1046	2 E83593	RND multidrug effl
25	38	52.1	3396	1 A42551	genome polypeptin
26	37	50.7	296	2 C95326	AttA2-like ABC tra
27	37	50.7	303	2 AE3520	sugar ABC transpor
28	37	50.7	514	1 VHNZCV	nucleocapsid prote
29	37	50.7	540	2 A11050	probable acyl Co-A

30	37	50.7	638	2 AB0446	probable exported
31	37	50.7	980	2 T20198	hypothetical prote
32	37	50.7	1108	2 D96798	hypothetical prote
33	36	49.3	106	2 S76260	hypothetical prote
34	36	49.3	163	2 G75435	hypothetical prote
35	36	49.3	179	2 T22143	hypothetical prote
36	36	49.3	197	1 R5MUL9	ribosomal protein
37	36	49.3	222	2 D82822	NADH-ubiquinone ox
38	36	49.3	288	2 A56279	carbon-monoxide de
39	36	49.3	334	2 S46805	hypothetical prote
40	36	49.3	518	2 D71288	probable glutamate
41	36	49.3	541	2 I41124	acyl CoA dehydroge
42	36	49.3	546	2 C91274	probable acyl coen
43	36	49.3	602	2 AB0024	probable potassium
44	36	49.3	643	2 B25817	nonstructural prot
45	36	49.3	876	2 E96674	hypothetical prote
46	36	49.3	876	2 E96674	genome polypeptin
47	36	49.3	3388	1 GNWVDP	genome polypeptin
48	36	49.3	3391	1 GNWV16	genome polypeptin
49	36	49.3	3391	1 GNWV26	polyprotein - deng
50	36	49.3	3391	2 JS0219	hypothetical prote
51	35	47.9	137	2 S37848	pro-neural chaete
52	35	47.9	199	2 A56548	hypothetical prote
53	35	47.9	202	2 E72688	hypothetical prote
54	35	47.9	241	2 D69205	hypothetical prote
55	35	47.9	261	2 T29076	conserved hypotet
56	35	47.9	271	2 H64702	hypothetical prote
57	35	47.9	271	2 G71817	hypothetical prote
58	35	47.9	310	2 G96986	sugar ABC transpor
59	35	47.9	310	2 E95406	probable ABC trans
60	35	47.9	314	2 PRLJSA	proteinase (EC 3.4
61	35	47.9	314	2 PRLJMP	probable two compo
62	35	47.9	420	2 H95966	sensor protein Cpx
63	35	47.9	464	2 C82046	hypothetical prote
64	35	47.9	468	2 T50873	cobyrinic acid synth
65	35	47.9	485	2 AG2921	cobyrinic acid synth
66	35	47.9	485	2 H97695	hypothetical prote
67	35	47.9	488	2 D70614	protein-tyrosine-p
68	35	47.9	1118	1 A49724	vitellogenin vit-2
69	35	47.9	1613	2 A43081	protein vit-2 [imp
70	35	47.9	1613	2 F89528	vitellogenin vit-1
71	35	47.9	1616	2 T16600	conserved hypotet
72	34.5	47.3	589	2 F86549	hypothetical prote
73	34.5	47.3	589	2 D72074	conserved hypotet
74	34	46.6	111	2 D72466	hypothetical prote
75	34	46.6	156	2 E83233	conserved hypotet
76	34	46.6	217	2 S74394	phosphoribosylglyc
77	34	46.6	257	2 D71544	hypothetical prote
78	34	46.6	275	2 A34866	T-cell surface pro
79	34	46.6	294	2 JC5000	aryl sulfotransfer
80	34	46.6	313	2 F86295	T24D18.18 protein
81	34	46.6	323	2 AD0994	probable membrane
82	34	46.6	342	1 T24660	probable phosphoes
83	34	46.6	414	2 F80186	probable 2,3-bisph
84	34	46.6	463	2 D87510	fumarate hydratase
85	34	46.6	470	2 T20851	hypothetical prote
86	34	46.6	491	2 C70868	hypothetical prote
87	34	46.6	505	1 S76229	DNA repair protein
88	34	46.6	510	2 T20850	hypothetical prote
89	34	46.6	534	2 T26238	hypothetical prote
90	34	46.6	598	2 C82194	ATP-dependent prot
91	34	46.6	720	2 A10467	DNA helicase II (E
92	34	46.6	801	1 D70309	ribonucleoside-dip
93	34	46.6	958	2 T13593	hypothetical prote
94	34	46.6	1011	2 C87789	protein C34G6.6 [i
95	34	46.6	1306	2 T13592	hypothetical prote
96	34	46.6	1516	2 F83085	conserved hypotet
97	34	46.6	3391	1 GNWVJA	genome polypeptin
98	33.5	45.9	181	2 F85568	negative modulator
99	33.5	45.9	181	2 F90718	negative modulator
100	33.5	45.9	181	2 A54296	seqA protein - Esc

## ALIGNMENTS

## RESULT 1

## HLHMSG

Class II histocompatibility antigen-associated gamma chain - human  
N:Alternate names: CD74 antigen; HLA-DR antigen-associated invariant chain; MHC class II  
C:Species: Homo sapiens (man)  
C>Date: 04-Dec-1986 #sequence revision 26-Jan-1996 #text change 31-Dec-2004  
C:Accession: A93981; A93981; A93602; A94103; S28903; S07182; A33234; A02243; A27551; A30  
R:Clasession: L.; Larhammar, D.; Raek, L.; Peterson, P.A.  
Proc. Natl. Acad. Sci. U.S.A. 80, 7395-7399, 1983  
A:Title: cDNA clone for the human invariant gamma chain of class II histocompatibility a  
A:Reference number: A93981; MUID:84170234; PMID:6324166  
A:Accession: A93981  
A:Molecule type: mRNA  
A:Residues: 1-216 <CL1>  
A:Cross-references: UNIPROT:P04233; UNIPROT:O19685; UNIPARC:UPI000016AA48; GB:K01144; NI  
A:Accession: B93981  
A:Molecule type: protein  
A:Residues: 1,'XXXXX',7,'XXXXXX',14,'X',16-17 <CL2>  
A:Cross-references: UNIPARC:UPI0000173828  
A:Note: radiochemical amino acid sequencing after cell-free translation  
R:Kudo, J.; Chao, L.Y.; Narni, F.; Saunders, G.F.  
Nucleic Acids Res. 13, 8827-8841, 1985  
A:Title: Structure of the human gene encoding the invariant gamma-chain of class II hist  
A:Reference number: A93602; MUID:86093681; PMID:3001652  
A:Accession: A93602  
A:Molecule type: DNA  
A:Residues: 'MHRRSRSCREDQKPV',1-150,'T',152-216 <KUD>  
A:Cross-references: UNIPARC:UPI0000173829  
A:Note: The authors translated the codon CAG for residue 109 as Gly, AAG for residue 143  
R:O'Sullivan, D.M.; Larhammar, D.; Wilson, M.C.; Peterson, P.A.; Quaranta, V.  
Proc. Natl. Acad. Sci. U.S.A. 83, 4484-4488, 1986  
A:Title: Structure of the human Ia-associated invariant (gamma)-chain gene: identificati  
A:Reference number: A94103; MUID:86233451; PMID:3459184  
A:Accession: A94103  
A:Molecule type: DNA  
A:Residues: 1-216 <OSU>  
A:Cross-references: UNIPARC:UPI000016AA48  
R:Riberdy, J.M.; Newcomb, J.R.; Surman, M.J.; Barbosa, J.A.; Cresswell, P.  
Nature 360, 474-477, 1992  
A:Title: HLA-DR molecules from an antigen-processing mutant cell line are associated wit  
A:Reference number: S28903; MUID:93078879; PMID:1448172  
A:Accession: S28903  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 81-104 <RIB>  
A:Cross-references: UNIPARC:UPI000002F1B4  
R:Strubin, M.; Mach, B.; Long, E.O.  
EMBO J. 3, 869-872, 1984  
A:Title: The complete sequence of the mRNA for the HLA-DR--associated invariant chain re  
A:Reference number: S07182; MUID:84207945; PMID:6586420  
A:Accession: S07182  
A:Molecule type: mRNA  
A:Residues: 'MHRRSRSCREDQKPV',1-216 <STR>  
A:Cross-references: UNIPARC:UPI000002AB09; EMBL:X00497; NID:G32130; PIDN:CAA25192.1; PID  
A:Accession: A33234  
A:Molecule type: DNA  
A:Residues: 'MHRRSRSCREDQKPV',1-4 <STR2>  
A:Cross-references: UNIPARC:UPI00000895C2  
R:Katunuma, N.; Kakegawa, H.; Matsunaga, Y.; Saibara, T.  
FEBS Lett. 349, 265-269, 1994  
A:Title: Immunological significances of invariant chain from the aspect of its structura  
A:Reference number: S46255; MUID:94326933; PMID:8050579  
A:Contents: annotation  
A:Note: some conclusions in this reference are based on the assumption, with no experime  
C:Comment: Class II antigens are associated with the invariant gamma chain during intrac  
a chains become independently integrated into the plasma membrane.  
C:Comment: Both cell-free translation experiments and the consensus sequence for the ini  
on of an additional sixteen residues upstream of the initiator, but this region is not l  
C:Genetics:  
A:Gene: GDB:CD74; DHLAG

A:Cross-references: GDB:119846; OMIM:142790

A:Map position: 5q31-5q33

A:Introns: 26/2, 84/1, 110/3, 131/3, 163/3, 193/1, 214/1

C:Superfamily: thyroglobulin type I repeat homology

C:Keywords: glycoprotein; transmembrane protein

F:1-30/Domain: intracellular #status predicted <INT>

F:31-56/Domain: transmembrane #status predicted <TM>

F:57-216/Domain: extracellular #status predicted <EXT>

F:114,120/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 73; DB 1; Length 216;

Best Local Similarity 100.0%; Pred. No. 2.7e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSQMRMATPLMQA 15

DB 87 PVSQMRMATPLMQA 101

## RESULT 2

## HLHMSG

Class II histocompatibility antigen-associated gamma chain Ii, 41k splice form - mouse  
N:Alternate names: Ia-associated invariant chain Ii41; invariant-chain proteoglycan core  
C:Contents: class II histocompatibility antigen-associated gamma chain Ii, 31k splice fo  
C:Species: Mus musculus (house mouse)

C>Date: 17-Mar-1987 #sequence revision 12-Apr-1996 #text change 09-Jul-2004

C:Accession: B27866; A27866; A02244; A43530; S03099; S03284; A28956

R:Koch, N.; Lauer, W.; Habicht, J.; Dobberstein, B.

EMBO J. 6, 1677-1683, 1987

A:Title: Primary structure of the gene for the murine Ia antigen-associated invariant ch  
nce of thyroglobulin.

A:Reference number: A91071; MUID:87275861; PMID:3038530

A:Accession: B27866

A:Molecule type: DNA

A:Residues: 1-279 <KOC>

A:Cross-references: UNIPROT:P04441; UNIPROT:O19452; UNIPARC:UPI000017382A

A:Experimental source: strain AKR

A:Note: the authors translated the codon AAC for residue 70 as Asp

A:Accession: A27866

A:Molecule type: DNA

A:Residues: 1-191,256-279 <KO2>

A:Cross-references: UNIPARC:UPI00000003E32

A:Experimental source: strain AKR

A:Note: the authors translated the codon AAC for residue 70 as Asp

R:Singler, P.A.; Lauer, W.; Dembic, Z.; Mayer, W.E.; Lipp, J.; Koch, N.; Hammerling, G.;

EMBO J. 3, 873-877, 1984

A:Title: Structure of the murine Ia-associated invariant (Ii) chain as deduced from a c

A:Reference number: A02244; MUID:84207946; PMID:6327293

A:Accession: A02244

A:Molecule type: mRNA

A:Residues: 47-191,256-279 <SIN>

A:Cross-references: UNIPARC:UPI000016CEFA; GB:X00496; NID:G53102; PIDN:CAA25191.1; PID:9

R:Eades, A.M.; Littfin, M.; Rahmsdorf, H.J.

J. Immunol. 144, 4399-4409, 1990

A:Title: The IFN-gamma response of the murine invariant chain gene is mediated by a com

A:Reference number: A43530; MUID:90257363; PMID:2111346

A:Accession: A43530

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-25 <END>

A:Cross-references: UNIPARC:UPI0000089341; GB:M35872; NID:G194129; PIDN:AAA37897.1; PID

R:Stone, J.; Perry, R.; Todd, J.A.; McDevitt, H.O.

submitted to the EMBL Data Library, March 1988

A:Reference number: S03099

A:Accession: S03099

A:Molecule type: mRNA

A:Residues: 1-58 <STO>

A:Cross-references: UNIPARC:UPI000016CEB3; EMBL:X07129; NID:G52637; PIDN:CAA30141.1; PID

R:Zhu, L.; Jones, P.P.

Nucleic Acids Res. 17, 447-448, 1989

A:Title: Complete sequence of the murine invariant chain (Ii) gene.

A:Reference number: S03284; MUID:89098406; PMID:2492095

A:Accession: S03284

A:Molecule type: DNA  
A:Residues: 1-9, 'M', '11-112, 'Q', '114-228, 'STG', '232-279 <ZHU>  
A:Cross-references: UNIPARC:UPI000017382B  
R:Miller, J.; Hatch, J.A.; Simonis, S.; Cullen, S.E.  
Proc. Natl. Acad. Sci. U.S.A. 85, 1359-1363, 1988  
A:Title: Identification of the glycosaminoglycan-attachment site of mouse invariant-chain  
A:Reference number: A28956; MUID:88144436; PMID:3422739  
A:Accession: A28956  
A:Molecule type: mRNA  
A:Residues: 261-270 <ML>  
A:Cross-references: UNIPARC:UPI000017382C  
A:Note: proof by site-directed mutagenesis that the sole chondroitin sulfate glycosaminoglycan  
C:Genetics:  
A:Introns: 26/2; 83/1; 109/3; 130/3; 161/3; 192/1; 256/1; 277/1  
C:Superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin  
C:Keywords: alternative splicing; glycoprotein; transmembrane protein  
F:1-279/Product: class II histocompatibility antigen-associated gamma chain II, 41K split  
F:1-191,256-279/Product: class II histocompatibility antigen-associated gamma chain II,  
F:1-29/Domin: intracellular #status predicted <CYT>  
F:30-55/Domin: transmembrane #status predicted <TM>  
F:56-279/Domin: extracellular #status predicted <EXT>  
F:196-254/Domin: thyroglobulin type I repeat homology <THY1>  
F:113,119,239,253/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:265/Binding site: carbohydrate (Ser) (covalent) #status experimental

Query Match 83.6%; Score 61; DB 1; Length 279;  
Best Local Similarity 85.7%; Pred. No. 0.0069;  
Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLLMO 14  
|||:|||||:  
Db 86 PVSKMRMATPLLMR 99

RESULT 3  
JC4796  
intracellular membrane glycoprotein type II invariant chain II - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 15-Oct-1995 #sequence\_revision 16-Aug-1996 #text\_change 09-Jul-2004  
C:Accession: JC4796  
R:Niimi, M.; Nakai, Y.; Aida, Y.  
Biochem. Biophys. Res. Commun. 222, 7-12, 1996  
A:Title: Identification of bovine invariant chain (II) gene by nucleotide sequencing.  
A:Reference number: JC4796; MUID:96212904; PMID:8630076  
A:Accession: JC4796  
A:Molecule type: mRNA  
A:Residues: 1-204 <NII>  
A:Cross-references: UNIPROT:Q29630; UNIPARC:UPI00008774C; DDBJ:D83962; NID:g1374680; PI  
C:Comment: This is a type II intracellular membrane glycoprotein. It binds major histoco  
s II expression and its function.  
C:Genetics:  
A:Gene: 11  
C:Superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin  
C:Keywords: glycoprotein; transmembrane protein  
F:1-30/Domin: intracellular (fragment) #status predicted <INT>  
F:12-15/Region: endosomal/lysosomal sorting signal  
F:31-56/Domin: transmembrane #status predicted <TM>  
F:57-204/Domin: extracellular #status predicted <EXT>  
F:81-104/Region: class II associated invariant chain CLIP motif  
F:112,118/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 82.2%; Score 60; DB 2; Length 204;  
Best Local Similarity 73.3%; Pred. No. 0.00076;  
Matches 11; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLLMOA 15  
|||:|||||:  
Db 87 PVSKMRMATPLLMLRA 101

RESULT 4  
S04362  
class II histocompatibility antigen-associated gamma chain, long splice form - rat

N:Contains: class II histocompatibility antigen-associated gamma chain, short splice form  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 09-Jul-2004  
R:McKnight, A.J.; Mason, D.W.; Barclay, A.N.  
C:Accession: S04362; S02182  
Nucleic Acids Res. 17, 3983-3984, 1989  
A:Title: Sequence of a rat MHC class II-associated invariant chain cDNA clone containing  
A:Reference number: S04362; MUID:89282409; PMID:2499873  
A:Accession: S04362  
A:Molecule type: mRNA  
A:Residues: 1-280 <MCK>  
A:Cross-references: UNIPROT:PI0247; UNIPARC:UPI000012C6BC; EMBL:X14254; NID:g56652; PIDN:  
R:Henkes, W.; Syha, J.; Reske, K.  
Nucleic Acids Res. 16, 11822, 1988  
A:Title: Nucleotide sequence of rat invariant gamma chain cDNA clone pUR-gamma-34.3.  
A:Reference number: S02182; MUID:89098337; PMID:3264906  
A:Accession: S02182  
A:Molecule type: mRNA  
A:Residues: 1-192,257-280 <HEN>  
A:Cross-references: UNIPARC:UPI00002AB0A; EMBL:X13044; NID:g56497; PIDN:CAA31450.1; PI  
A:Note: the authors translated the codon ATC for residue 16 as Asn, CTG for residue 17  
C:Superfamily: class II histocompatibility antigen-associated gamma chain; thyroglobulin  
C:Keywords: alternative splicing; transmembrane protein  
F:197-255/Domin: thyroglobulin type I repeat homology <THY1>

Query Match 80.8%; Score 59; DB 2; Length 280;  
Best Local Similarity 85.7%; Pred. No. 0.0017;  
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLLMO 14  
|||:|||||:  
Db 87 PVSKMRMATPLLMR 100

RESULT 5  
G84420  
TATC-like protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 24-Aug-2001  
C:Accession: G84420  
R:Lidin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: G84420  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-327 <STO>  
A:Cross-references: UNIPARC:UPI00000502CC; GB:AE002093; NID:g6598626; PIDN:AAF18659.1;  
C:Genetics:  
A:Gene: At2g0110  
A:Map position: 2  
C:Superfamily: conserved hypothetical protein HI0188

Query Match 57.5%; Score 42; DB 2; Length 327;  
Best Local Similarity 66.7%; Pred. No. 3.4;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLL 12  
|||:|||||:  
Db 301 PVTQMLLATPLL 312

RESULT 6  
T44583  
conserved hypothetical protein tatC [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T44583  
R:Summer, E.J.; McCaffery, M.W.; Cline, K.  
submitted to the EMBL Data Library, April 1999

A;Description: An Arabidopsis homolog of TatC/YCF43.  
A;Reference number: Z22800  
A;Accession: T44563  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-340 <SUM>  
A;Cross-references: UNIPROT:Q9SJVS; UNIPARC:UPI000003CA3E; EMBL:AF145045; PIDN:AAD33946.  
C;Superfamily: conserved hypothetical protein H10188

Query Match 57.5%; Score 42; DB 2; Length 340;  
Best Local Similarity 66.7%; Pred. No. 2;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKRMATPL 12  
||:|:|:|:|:|  
DB 314 PVTQMLLATPLL 325

## RESULT 7

T03471  
probable poly(3-hydroxyalkanoate) polymerase - Rhodobacter capsulatus  
C;Species: Rhodobacter capsulatus  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C;Accession: T03471  
R;Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.  
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997  
A;Title: Sequence of a 189-kb segment of the chromosomal of Rhodobacter capsulatus SB1003  
A;Reference number: Z14955; MUID:97404404; PMID:9256491  
A;Accession: T03471  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-92 <VLC>  
A;Cross-references: UNIPROT:O68038; UNIPARC:UPI00001318ED; EMBL:AF010496; NID:g3128256;  
C;Genetics:  
A;Map position: 1

Query Match 54.8%; Score 40; DB 2; Length 92;  
Best Local Similarity 57.1%; Pred. No. 2.2;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKRMATPLMQA 15  
|:|:|:|:|:|  
DB 53 VTGMRMGTPFLFEA 66

## RESULT 8

T00513  
cytochrome P450 homolog At2g23190 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004  
C;Accession: T00513; F84621  
R;Rounstey, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, November 1997  
A;Description: Arabidopsis thaliana chromosome II BAC T20D16 genomic sequence.  
A;Reference number: Z14164  
A;Accession: T00513  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-543 <ROU>  
A;Cross-references: UNIPROT:O22188; UNIPARC:UPI00000A8D78; EMBL:AC002391; NID:g2642427;  
A;Experimental source: cultivar Columbia  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: F84621  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-543 <STO>  
A;Cross-references: UNIPARC:UPI00000A8D78; GB:AE002093; NID:g2642444; PIDN:AAB87112.1; C  
C;Genetics:

A;Gene: T20D16.18; At2g23190

A;Map position: 2  
A;Introns: 211/3; 339/3  
C;Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C;Keywords: heme; iron; metalloprotein  
F;341-500/Domain: cytochrome P450 homology <P45>  
F;478/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 54.8%; Score 40; DB 2; Length 543;  
Best Local Similarity 66.7%; Pred. No. 14;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVSKRMATPLL 12  
||:|:|:|:|:|  
DB 515 PWMMRPATPLL 526

## RESULT 9

AG0918  
DNA helicase II [imported] - Salmonella enterica subsp. enterica serovar Typhi (strain C)  
C;Species: Salmonella enterica subsp. enterica serovar Typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AG0918  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AG0918  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-720 <PAR>  
A;Cross-references: UNIPARC:UPI000005A572; GB:AL513382; PIDN:CAD07941.1; PID:g16504486;  
C;Genetics:  
A;Gene: STY3608  
C;Superfamily: helicase II

Query Match 54.8%; Score 40; DB 2; Length 720;  
Best Local Similarity 72.7%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PVSKRMATPL 11  
||:|:|:|:|:|  
DB 653 PVSHQRMGTPL 663

## RESULT 10

T43466  
hypothetical protein DKFZp434P1723.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C;Accession: T43466  
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, December 1999  
A;Reference number: Z22517  
A;Accession: T43466  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-554 <AAA>  
A;Cross-references: UNIPROT:Q9UNN5; UNIPARC:UPI000016AC9D; EMBL:AL133631  
A;Experimental source: adult testis; clone DKFZp434P1723  
C;Genetics:  
A;Note: DKFZp434P1723.1

Query Match 53.4%; Score 39; DB 2; Length 554;  
Best Local Similarity 70.0%; Pred. No. 22;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKRMATPL 10  
||:|:|:|:|:|

Db 476 PVSKLRIRTP 485

RESULT 11

JC7093

Fas associated factor 1 - human

C:Species: Homo sapiens (man)

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: JC7093

R:Ryu, S.W.; Chae, S.K.; Lee, K.J.; Kim, E.

Biochem. Biophys. Res. Commun. 262, 388-394, 1999

A:Title: Identification and characterization of human Fas associated factor 1, hFAF1.

A:Reference number: JC7093; MUID:99393315; PMID:10462485

A:Accession: JC7093

A:Molecule type: mRNA

A:Residues: 1-650 <RYU>

A:Cross-references: UNIPROT:Q9UNN5; UNIPARC:UPI00001698CA; GB:AF106798; NID:G5805207; P1

A:Experimental source: Hela cell

Query Match 53.4%; Score 39; DB 2; Length 650;

Best Local Similarity 70.0%; Pred. No. 26;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PVSKRMATP 10

Db 572 PVSKLRIRTP 581

RESULT 12

HJEC22

DNA helicase II (EC 3.6.1.-) [validated] - Escherichia coli (strain K-12)

N:Contains: adenosinetriphosphatase (EC 3.6.1.3)

C:Species: Escherichia coli

C>Date: 17-Mar-1987 #sequence\_revision 10-Oct-1997 #text\_change 09-Jul-2004

C:Accession: F65185; J50014; A93528; A93498; S30703; E37841; A03549

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F65185

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-720 <BLAT>

A:Cross-references: UNIPROT:P03018; UNIPARC:UPI0000137EPA; GB:AE0000457; GB:U000096; NID:9

A:Experimental source: strain K-12, substrain MG1655

R:Yamamoto, Y.; Ogawa, T.; Shinagawa, H.; Nakayama, T.; Matsuo, H.; Ogawa, H.

J. Biochem. 99, 1579-1590, 1986

A:Title: Determination of the initiation sites of transcription and translation of the u

A:Reference number: A92004; MUID:86304220; PMID:2943729

A:Accession: J50014

A:Molecule type: DNA

A:Residues: 1-437, 'R', 439-720 <YAM>

A:Cross-references: UNIPARC:UPI000016F607; GB:D00069; NID:g216672; PIDN:BA000

A:Note: residues 1-24 were confirmed by protein sequencing

R:Finch, P.W.; Emerson, P.T.

Nucleic Acids Res. 12, 5789-5799, 1984

A:Title: The nucleotide sequence of the uvrD gene of Escherichia coli.

A:Reference number: A93528; MUID:84272253; PMID:6379604

A:Accession: A93528

A:Molecule type: DNA

A:Residues: 1-223, 'N', 225-290, 'N', 292, 'R', 294-329, 'T', 331-539, 'V', 541-720 <FIN>

A:Cross-references: UNIPARC:UPI000016F606; GB:X00738; GB:K01148; GB:X00225; NID:g43296;

R:Easton, A.M.; Kushner, S.R.

Nucleic Acids Res. 11, 8625-8640, 1983

A:Title: Transcription of the uvrD gene of Escherichia coli is controlled by the lexA re

A:Reference number: A93498; MUID:84169504; PMID:6324092

A:Accession: A93498

A:Molecule type: DNA

A:Residues: 1-258 <EAS>

A:Cross-references: UNIPARC:UPI000016F605; GB:X00225; NID:g43294; PIDN:CAA25043.1; PID:9

R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.

Science 257, 771-778, 1992

A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5

A:Reference number: S30660; MUID:92358234; PMID:1379743

A:Accession: S30703

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-47, 'X', 49-91, 'X', 93-97, 'XX', 100-720 <DAN>

A:Cross-references: UNIPARC:UPI0000172EB9; EMBL:M87049

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1992

R:Colloms, S.D.; Sykora, P.; Szatmari, G.; Sherratt, D.J.

J. Bacteriol. 172, 6973-6980, 1990

A:Title: Recombination at ColE1 cer requires the Escherichia coli xerC gene product, a

A:Reference number: A37841; MUID:91072248; PMID:2254268

A:Accession: E37841

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-17 <COL>

A:Cross-references: UNIPARC:UPI0000172EBA; GB:M38257

C:Comment: The gene that codes for this protein is induced during the SOS response; the

C:Genetics:

A:Gene: uvrD

A:Map position: 86 min

A:Note: expression is induced by DNA-damaging agents, e.g. nalidixic acid or mitomycin

C:Function: <HEL>

C:Function: shows ATP-dependent DNA-unwinding activity [validated, MUID:84111621]

C:Description: adenosinetriphosphatase (EC 3.6.1.3), single-stranded DNA-dependent [val

C:Superfamily: helicase II

C:Keywords: ATP; DNA binding; DNA repair; DNA replication; hydrolase; nucleotide binding

F:29-36/Region: nucleotide-binding motif A (P-loop)

Query Match 53.4%; Score 39; DB 1; Length 720;

Best Local Similarity 58.3%; Pred. No. 29;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PVSKRMATPLL 12

Db 653 PVSHQRMGTPMV 664

RESULT 13

G91221

DNA-dependent ATPase I and helicase II [imported] - Escherichia coli (strain O157:H7, s

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C:Accession: G91221

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: G91221

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-720 <HAY>

A:Cross-references: UNIPROT:O8X8P5; UNIPARC:UPI000000D08F3; GB:BA000007; PIDN:BA038166.1

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs4743

C:Superfamily: helicase II

Query Match 53.4%; Score 39; DB 2; Length 720;

Best Local Similarity 58.3%; Pred. No. 29;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PVSKRMATPLL 12

Db 653 PVSHQRMGTPMV 664

RESULT 14

B6068

DNA-dependent ATPase I and helicase II [imported] - Escherichia coli (strain O157:H7, s

C:Species: Escherichia coli







F;472-488/Domain: transmembrane #status predicted <TM6>  
F;540-556/Domain: transmembrane #status predicted <TM7>  
F;874-890/Domain: transmembrane #status predicted <TM8>  
F;898-914/Domain: transmembrane #status predicted <TM9>  
F;974-990/Domain: transmembrane #status predicted <TM10>  
F;1012-1028/Domain: transmembrane #status predicted <TM11>

Query Match 52.1%; Score 38; DB 2; Length 1034;  
Best Local Similarity 58.3%; Pred. No. 67;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 SKMRMATPLLQ 14  
:|:::|||||  
Db 109 NKQLATPLLPQ 120

RESULT 23  
S39630

multidrug-efflux transport protein B - Pseudomonas aeruginosa

N;Alternate names: multidrug resistance protein B

C;Species: Pseudomonas aeruginosa

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004

C;Accession: S39630

R;Poole, K.; Heinrichs, D.E.; Neshat, S.

Mol. Microbiol. 10, 529-544, 1993  
A;Title: Cloning and sequence analysis of an EnvCD homologue in Pseudomonas aeruginosa:

A;Reference number: S39629; MUID:95058196; PMID:7968531

A;Accession: S39630

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1046 <POO>

A;Cross-references: UNIPROT:P52002; UNIPARC:UPI000016FC9A; GB:L111616; NID:g438852; PIDN:

A;Experimental source: strain CD10

C;Genetics:

A;Gene: mexB

C;Function:

A;Description: probably involved in secretion of the siderophore pyoverdine

A;Note: inducible under conditions of iron limitation

C;Superfamily: acriflavin resistance protein

C;Keywords: transmembrane protein; transport protein

F;14-30/Domain: transmembrane #status predicted <TM1>

F;343-359/Domain: transmembrane #status predicted <TM2>

F;370-386/Domain: transmembrane #status predicted <TM3>

F;397-413/Domain: transmembrane #status predicted <TM4>

F;442-458/Domain: transmembrane #status predicted <TM5>

F;472-488/Domain: transmembrane #status predicted <TM6>

F;541-557/Domain: transmembrane #status predicted <TM7>

F;874-900/Domain: transmembrane #status predicted <TM8>

F;898-914/Domain: transmembrane #status predicted <TM9>

F;973-989/Domain: transmembrane #status predicted <TM10>

F;1014-1030/Domain: transmembrane #status predicted <TM11>

Query Match 52.1%; Score 38; DB 2; Length 1046;  
Best Local Similarity 58.3%; Pred. No. 67;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 SKMRMATPLLQ 14  
:|:::|||||  
Db 109 NKQLATPLLPQ 120

RESULT 24

E83593

RND multidrug efflux transporter MexB PA0426 [imported] - Pseudomonas aeruginosa (strain

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004

C;Accession: E83593

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: E83593

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1046 <STO>

A;Cross-references: UNIPROT:P52002; UNIPARC:UPI000012F022; GB:AE004479; GB:AE004091; NID:

A;Experimental source: strain PA01

C;Genetics:

A;Gene: mexB; PA0426

C;Superfamily: acriflavin resistance protein

Query Match 52.1%; Score 38; DB 2; Length 1046;  
Best Local Similarity 58.3%; Pred. No. 67;  
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 SKMRMATPLLQ 14  
:|:::|||||  
Db 109 NKQLATPLLPQ 120

RESULT 25

A42551

genome polyprotein - dengue virus type 1 (strain Singapore S275/90)

N;Contents: capsid protein; envelope protein; membrane protein; nonstructural protein NS5

a; nonstructural protein NS4b; nonstructural protein NS5

C;Species: dengue virus type 1

C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004

C;Accession: A42551

R;Fu, J.; Tan, B.H.; Yap, E.H.; Chan, Y.C.; Tan, Y.H.

Virology 188, 953-958, 1992

A;Title: Full-length cDNA sequence of dengue type 1 virus (Singapore strain S275/90).

A;Reference number: A42551; MUID:92263809; PMID:1585663

A;Accession: A42551

A;Molecule type: genomic RNA

A;Residues: 1-3396 <FOU>

A;Cross-references: UNIPROT:P33478; UNIPARC:UPI000002F845; GB:M87512

C;Superfamily: hepatitis C virus genome polyprotein

C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;

F;1-114/Product: capsid protein #status predicted <CAP>

F;115-281/Product: membrane protein precursor #status predicted <MEP>

F;115-204/Domain: nonterminal signal sequence #status predicted <SIG>

F;205-281/Product: membrane protein #status predicted <MEM>

F;267-279/Domain: transmembrane #status predicted <TM1>

F;282-774/Product: envelope protein #status predicted <ENV>

F;753-769/Domain: transmembrane #status predicted <TM2>

F;773-1127/Product: nonstructural protein NS1 #status predicted <NS1>

F;1128-1344/Product: nonstructural protein NS2a #status predicted <N2a>

F;1345-1474/Product: nonstructural protein NS2b #status predicted <N2b>

F;1475-2093/Product: nonstructural protein NS3 #status predicted <NS3>

F;1668-1675/Region: nucleotide-binding motif A (P-loop)

F;1755-1760/Region: nucleotide-binding motif B

F;1759-1762/Region: DEAH motif

F;2094-2243/Product: nonstructural protein NS4a #status predicted <N4a>

F;2244-2492/Product: nonstructural protein NS4b #status predicted <N4b>

F;2493-3396/Product: nonstructural protein NS5 #status predicted <NS5>

F;183,347,433/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.1%; Score 38; DB 1; Length 3396;  
Best Local Similarity 58.3%; Pred. No. 2.3e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PVSKMRMATPLL 12  
:|:::|||||  
Db 2327 PISKMDIGVPLL 2338

RESULT 26

C95326

ActA2-like ABC transporter, permease protein SMA0952 [imported] - Sinorhizobium meliloti

C;Species: Sinorhizobium meliloti

C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004

C;Accession: C95326

R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bow

.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
A:Reference number: A95262; MUID:21396509; PMID:11481432  
A:Accession: C95326  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-296 <KUR>  
A:Cross-references: UNIPROT:Q92ZHO; UNIPARC:UPI00000CB0FC; GB:AE006469; PIDN:AAK65173.1;  
A:Experimental source: strain 1021, megaplasmid pSynA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chaint, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMA0952  
A:Genome: plasmid  
C:Superfamily: spermidine/putrescine transport system permease protein poth

Query Match 50.7%; Score 37; DB 2; Length 296;  
Best Local Similarity 50.0%; Pred. No. 27;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLL 12  
:::|||||  
Db 25 PIASMAVTPPLL 36

RESULT 27  
AE3520  
sugar ABC transporter, permease protein BMEII0087 [imported] - Brucella melitensis (stra  
C:Species: Brucella melitensis  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
C:Accession: AE3520  
R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, .; Mazur, M.; Goltzman, E.; Selkov, E.; Elzeir, P.H.; Hagijs, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AE3520  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-303 <KUR>  
A:Cross-references: UNIPROT:Q8YDT8; UNIPARC:UPI0000058346; GB:AE008918;  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEII0087  
A:Map position: II  
C:Superfamily: probable ribose ABC transporter rbsC-2

Query Match 50.7%; Score 37; DB 2; Length 303;  
Best Local Similarity 63.6%; Pred. No. 28;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 5 MRMATPLLMOA 15  
:::|||||  
Db 18 LRVATPLLAA 28

RESULT 28  
VHNZCV  
nucleocapsid protein - canine distemper virus (fragment)  
C:Species: canine distemper virus  
C:Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 14-Nov-1997  
C:Accession: A04028  
R:Rozenblatt, S.; Eizenberg, O.; Ben-Levy, R.; Lavie, V.; Bellini, W.J.  
J. Virol. 53, 684-690, 1985  
A:Title: Sequence homology within the morbilliviruses.  
A:Reference number: A93006; MUID:85108164; PMID:3838193

A:Accession: A04028  
A:Molecule type: Genomic RNA  
A:Residues: 1-514 <ROZ>  
A:Cross-references: UNIPARC:UPI0000174A76; GB:M10242; NID:G323245  
A:Note: the codons given for residues 54-Lys (AAU) and 55-Ile (GUC) are inconsistent with  
C:Superfamily: paramyxovirus nucleocapsid protein  
C:Keywords: nucleocapsid

Query Match 50.7%; Score 37; DB 1; Length 514;  
Best Local Similarity 53.8%; Pred. No. 49;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 VSKMRMATPLLMO 14  
:::|||||  
Db 484 IARMLRMLTKMLSQ 496

RESULT 29  
AII050  
probable acyl Co-A dehydrogenase [imported] - Salmonella enterica subsp. enterica serov  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AII050  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AII050  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <PAR>  
A:Cross-references: UNIPARC:UPI00000CDBCE; GB:AL513382; PIDN:CAD06854.1; PID:gl6505502;  
C:Genetics:  
A:Gene: aidB

Query Match 50.7%; Score 37; DB 2; Length 540;  
Best Local Similarity 57.1%; Pred. No. 52;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLLMO 14  
:::|||||  
Db 134 PVMTFAATPLLQ 147

RESULT 30  
AB0446  
probable exported protein YPO3664 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 31-Dec-2004  
C:Accession: AB0446  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AB0446  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-638 <KUR>  
A:Cross-references: UNIPROT:Q8ZAW7; UNIPARC:UPI00000DCE73; GB:AL590842; PIDN:CAC93134.1  
C:Genetics:  
A:Gene: YPO3664  
C:Superfamily: MSHA biogenesis protein MshH

Query Match 50.7%; Score 37; DB 2; Length 638;  
Best Local Similarity 50.0%; Pred. No. 62;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PVSKRMATPLLMQ 14  
|::| |::| |::|  
Db 57 PINKREWPLPMQ 70

## RESULT 31

T20198  
hypotheical protein M01F1.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20198; T23659  
R:Lightning, J.  
submitted to the EMBL Data Library, July 1996  
A:Reference number: Z19236  
A:Accession: T20198  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-980 <W1>  
A:Cross-references: UNIPROT:O62119; UNIPARC:UPI00000784BB; EMBL:Z77131; PIDN:CAB00857.1;  
A:Experimental source: clone C54C6  
R:Sim, M.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: Z19778  
A:Accession: T23659  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-980 <W12>  
A:Cross-references: UNIPARC:UPI00000784BB; EMBL:Z46381; PIDN:CAA86520.1; GSPDB:GN00021;  
A:Experimental source: clone M01F1  
C:Genetics:  
A:Gene: CESP:M01F1.7  
A:Map position: 3  
A:Introns: 26/3; 71/1; 234/3; 375/2; 506/3; 539/1; 580/2; 639/2; 687/3; 927/1; 958/2

Query Match 50.7%; Score 37; DB 2; Length 980;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVSKRMATPLL 12  
|::| |::| |::|  
Db 93 PVTKRYSTPMW 104

## RESULT 32

D96798  
hypotheical protein F22K20.5 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: D96798  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D96798  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1108 <STO>  
A:Cross-references: UNIPROT:O49281; UNIPARC:UPI00000A0F07; GB:AE005173; NID:G2829910; PT  
C:Genetics:  
A:Gene: F22K20.5  
A:Map position: 1

Query Match 50.7%; Score 37; DB 2; Length 1108;  
Best Local Similarity 57.1%; Pred. No. 1.le+02;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 VSKRMATPLLMQA 15  
|::| |::| |::|  
Db 782 IGMRRATPKLAQA 795

## RESULT 33

S76260  
hypotheical protein slr0770 - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: S76260  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S76260  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-106 <KAN>  
A:Cross-references: UNIPROT:Q55605; UNIPARC:UPI00000C0F10; EMBL:D64000; GB:AB001339; NID:  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 49.3%; Score 36; DB 2; Length 106;  
Best Local Similarity 58.3%; Pred. No. 14;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 KMRMATPLLMQA 15  
|::| |::| |::|  
Db 85 EMRSSTPSLMEA 96

## RESULT 34

G75435  
hypotheical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: G75435  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Me  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: G75435  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-163 <WHI>  
A:Cross-references: UNIPROT:Q9RVB7; UNIPARC:UPI00000C1897; GB:AE001961; GB:AE000513; NID:  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR112  
A:Map position: 1

Query Match 49.3%; Score 36; DB 2; Length 163;  
Best Local Similarity 53.8%; Pred. No. 23;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 PVSKRMATPLLM 13  
|::| |::| |::|  
Db 11 PFSHVLRLAGPLLL 23

## RESULT 35

T22143  
hypotheical protein F43G9.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T22143  
R:Kershaw, J.  
submitted to the EMBL Data Library, September 1996



A;Experimental source: strain OM5, plasmid pHC3  
A;Note: sequence extracted from NCBI backbone (NCBIP:111732); sequence misidentified as C;Genetics:

A;Gene: coxM; codH

A;Start codon: GTG

C;Complex: heterotrimer of large (see PIR:C56279), medium, and small (see PIR:B56279) chain

C;Superfamily: carbon monoxide dehydrogenase medium chain

C;Keywords: heterotrimer; oxidoreductase

Query Match 49.3%; Score 36; DB 2; Length 288;

Best Local Similarity 60.0%; Pred. No. 41;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PVSKRMATP 10

| : | : | : |

Db 39 PINKTRLATP 48

#### RESULT 39

S46805

hypothetical protein YHR085w - Yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 09-Jul-2004

C;Accession: S46805

R;Favella, T.

submitted to the EMBL Data Library, June 1994

A;Description: The sequence of S. cerevisiae cosmid 9205.

A;Reference number: S46795

A;Accession: S46805

A;Molecule type: DNA

A;Residues: 1-334 <FAV>

A;Cross-references: UNIPROT:P38803; UNIPARC:UPI0000053233; EMBL:U10556; NID:G500825; PID

C;Genetics:

A;Gene: MIPS:YHR085w

A;Cross-references: SGD:S0001127

A;Map position: 8R

Query Match 49.3%; Score 36; DB 2; Length 334;

Best Local Similarity 50.0%; Pred. No. 48;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSKRMATPLMQA 15

| : | : | : |

Db 84 IIKSRLMTPLLTQS 97

#### RESULT 40

D71288

probable glutamate synthase (gluA) - syphilis spirochete

C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Oct-2004

C;Accession: D71288

R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A;Reference number: A71250; PMID:98332770; PMID:9665876

A;Accession: D71288

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-518 <COL>

A;Cross-references: UNIPROT:O83717; UNIPARC:UPI000003295; GB:AE001245; GB:AE000520; NID

A;Experimental source: strain Nichols

C;Genetics:

A;Gene: TP0735

C;Superfamily: glutamate synthase, small subunit

Query Match 49.3%; Score 36; DB 2; Length 518;

Best Local Similarity 50.0%; Pred. No. 77;

Matches 9; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 1 PVSKRMATP----LLMQ 14

Db 58 PLQRMMLPPKKRSLLMQ 75

| : | : | : |

Search completed: September 9, 2006, 23:01:55

Job time : 20.0886 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: September 9, 2006, 22:38:21 ; Search time 137.468 Seconds  
(without alignments)  
100.934 Million cell updates/sec

Title: US-10-617-568-1

Perfect score: 73

Sequence: 1 PVSKRMATPLLMOA 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : UniProt 7.2.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	100.0	25	2	Q9TNQ6 HUMAN
2	73	100.0	160	2	O8SNA0 homo sapien
3	73	100.0	232	2	Q5U0J8 HUMAN
4	73	100.0	232	2	Q5U0J8 homo sapien
5	73	100.0	296	1	Q5RFJ4 pongo pygma
6	69	94.5	20	2	P04233 homo sapien
7	63	86.3	208	2	Q9TNQ4 HUMAN
8	62	84.9	208	2	Q9TNQ4 homo sapien
9	61	83.6	215	2	Q9MXD5 HORSE
10	61	83.6	279	1	O8K0S8 MOUSE
11	61	83.6	279	1	Q8GT22 MOUSE
12	60	82.2	190	2	P04441 mus musculus
13	60	82.2	204	2	Q3U4Q8 mouse
14	59	80.8	216	2	Q7JFY1 BOVIN
15	59	80.8	280	1	Q29630 BOVIN
16	56	76.7	214	2	Q6GT70 RAT
17	42	57.5	340	2	HG2A RAT
18	42	57.5	1054	2	Q764N1 PIG
19	41	56.2	110	2	Q9SJV5 ARATH
20	41	56.2	238	1	Q3G4U7 9DELTA
21	41	56.2	471	2	Q3AZT7 SYN9
22	40	54.8	92	1	Q3YUFI NOCFA
23	40	54.8	335	2	Q5YUFI nocardia fa
24	40	54.8	350	2	O68038 rhodobacter
25	40	54.8	335	2	Q7Q9V4 ANOHA
26	40	54.8	675	2	Q7Q9V4 anopheles g
27	40	54.8	720	1	O22188 ARATH
28	40	54.8	720	2	Q61DK9 CAEBR
29	40	54.8	720	2	Q57HQ6 SALCH
30	40	54.8	720	2	O5PKM2 SALCH
31	40	54.8	755	2	Q823B0 SALT
					Q21RQ8 RHODOPSEUDO

32	39	53.4	140	2	Q9BZT1_HUMAN	Q9bzt1 homo sapien
33	39	53.4	141	2	Q47S32_THERFY	Q47s32 thermobifid
34	39	53.4	304	2	Q18767_CAEEL	Q18767 caenorhabdi
35	39	53.4	411	2	Q8TP38_METAC	Q8tp38 metanosarc
36	39	53.4	464	2	Q5VTAL_HUMAN	Q5vtal homo sapien
37	39	53.4	469	2	Q2U394_ASPOR	Q2u394 aspergillus
38	39	53.4	490	2	Q5VTA2_HUMAN	Q5vta2 homo sapien
39	39	53.4	505	2	Q9FY6_CAPAN	Q9fy6 capsicum an
40	39	53.4	524	2	Q6CSN2_YARLI	Q6csn2 yarrowia li
41	39	53.4	589	2	Q5WJ2_ORYSA	Q5wj2 oryza sativ
42	39	53.4	649	1	FAF1_MOUSE	P54731 mus musculu
43	39	53.4	649	1	FAF1_RAT	Q924k2 rattus norv
44	39	53.4	649	2	Q6P1F8_MOUSE	Q6p1f8 mus musculu
45	39	53.4	650	1	FAF1_HUMAN	Q9unn5 homo sapien
46	39	53.4	650	2	Q5VTA3_HUMAN	Q5vta3 homo sapien
47	39	53.4	670	2	Q8AAK8_BACTN	Q8aak8 bacteroides
48	39	53.4	720	1	UVRD_ECOLI	P30108 escherichia
49	39	53.4	720	2	Q3IUH5_SHIBS	Q3iuh5 shigella bo
50	39	53.4	720	2	Q329Y9_SHIDS	Q329y9 shigella dy
51	39	53.4	720	2	Q3YVF3_SHISS	Q3yvf3 shigella so
52	39	53.4	720	2	Q2M8B9_ECOLI	Q2m8b9 escherichia
53	39	53.4	720	2	O8X8P5_ECO57	Q8x8p5 escherichia
54	39	53.4	720	2	O8J1W7_SHIFL	Q8j1w7 shigella fl
55	39	53.4	738	2	Q8FBN3_ECOL6	Q8fbn3 escherichia
56	39	53.4	769	2	Q6CP89_KUULA	Q6cp89 kluyveromyc
57	39	53.4	1287	2	Q2QVE1_ORYSA	Q2qve1 oryza sativ
58	39	53.4	1306	2	Q37170_DEN3	Q37170 dengue viru
59	39	53.4	3311	2	Q6U7P9_9FLAV	Q6u7p9 dengue viru
60	39	53.4	3390	1	POLG_DEN3	P27915 d genome po
61	39	53.4	3390	2	Q2TN90_DEN3	Q2tn90 dengue viru
62	39	53.4	3390	2	Q2VM87_DEN3	Q2vm87 dengue viru
63	39	53.4	3390	2	Q2VM88_DEN3	Q2vm88 dengue viru
64	39	53.4	3390	2	Q2VM89_DEN3	Q2vm89 dengue viru
65	39	53.4	3390	2	Q2VM90_DEN3	Q2vm90 dengue viru
66	39	53.4	3390	2	Q2VM91_DEN3	Q2vm91 dengue viru
67	39	53.4	3390	2	Q2VM92_DEN3	Q2vm92 dengue viru
68	39	53.4	3390	2	Q2VM96_DEN3	Q2vm96 dengue viru
69	39	53.4	3390	2	Q2MG69_DEN3	Q2mg69 dengue viru
70	39	53.4	3390	2	Q2MG70_DEN3	Q2mg70 dengue viru
71	39	53.4	3390	2	Q2MG71_DEN3	Q2mg71 dengue viru
72	39	53.4	3390	2	Q3L2Z4_DEN3	Q3l2z4 dengue viru
73	39	53.4	3390	2	Q3L2Z5_DEN3	Q3l2z5 dengue viru
74	39	53.4	3390	2	Q3L2Z6_DEN3	Q3l2z6 dengue viru
75	39	53.4	3390	2	Q3L2Z7_DEN3	Q3l2z7 dengue viru
76	39	53.4	3390	2	Q3L2Z8_DEN3	Q3l2z8 dengue viru
77	39	53.4	3390	2	Q3L2Z9_DEN3	Q3l2z9 dengue viru
78	39	53.4	3390	2	Q3L300_DEN3	Q3l300 dengue viru
79	39	53.4	3390	2	Q3L301_DEN3	Q3l301 dengue viru
80	39	53.4	3390	2	Q3L302_DEN3	Q3l302 dengue viru
81	39	53.4	3390	2	Q5D0W9_DEN3	Q5d0w9 dengue viru
82	39	53.4	3390	2	Q5FY70_DEN3	Q5fy70 dengue viru
83	39	53.4	3390	2	Q5I3B8_DEN3	Q5i3b8 dengue viru
84	39	53.4	3390	2	Q5I3B9_DEN3	Q5i3b9 dengue viru
85	39	53.4	3390	2	Q5I3C0_DEN3	Q5i3c0 dengue viru
86	39	53.4	3390	2	Q5I3C1_DEN3	Q5i3c1 dengue viru
87	39	53.4	3390	2	Q5I3C2_DEN3	Q5i3c2 dengue viru
88	39	53.4	3390	2	Q5I3C3_DEN3	Q5i3c3 dengue viru
89	39	53.4	3390	2	Q5I3C4_DEN3	Q5i3c4 dengue viru
90	39	53.4	3390	2	Q5I3C5_DEN3	Q5i3c5 dengue viru
91	39	53.4	3390	2	Q5I3C6_DEN3	Q5i3c6 dengue viru
92	39	53.4	3390	2	Q5I3C7_DEN3	Q5i3c7 dengue viru
93	39	53.4	3390	2	Q5I3C8_DEN3	Q5i3c8 dengue viru
94	39	53.4	3390	2	Q5I3C9_DEN3	Q5i3c9 dengue viru
95	39	53.4	3390	2	Q5UB51_DEN3	Q5ub51 dengue viru
96	39	53.4	3390	2	Q689F5_DEN3	Q689f5 dengue viru
97	39	53.4	3390	2	Q689F6_DEN3	Q689f6 dengue viru
98	39	53.4	3390	2	Q689F7_DEN3	Q689f7 dengue viru
99	39	53.4	3390	2	Q689F8_DEN3	Q689f8 dengue viru
100	39	53.4	3390	2	Q6B523_DEN3	Q6b523 dengue viru

ALIGNMENTS

```

RESULT 1
Q9TNQ6 HUMAN PRELIMINARY; PRT; 25 AA.
AC Q9TNQ6_1
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
DT 01-MAY-2000, sequence version 1.
DE 07-FEB-2006, entry version 10.
DE HLA class II-associated invariant chain II (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP PROTEIN SEQUENCE.
RX MEDLINE=92375195; PubMed=1380674; DOI=10.1038/358764a0;
RA Chicz R.M., Urban R.G., Lane W.S., Gorga J.C., Stern L.J.,
RA Vignali D.A., Strominger J.L.;
RT "Predominant naturally processed peptides bound to HLA-DR1 are derived
RT from MHC-related molecules and are heterogeneous in size.";
RL Nature 358:764-768(1992).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR Ensembl; ENSG00000019582; Homo sapiens.
FT NON_TER 1 1
FT NON_TER 25 25
SQ SEQUENCE 25 AA; 2793 MW; 2F80100276A802F4 CRC64;

Query Match 100.0%; Score 73; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 2.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLMQA 15
Db 7 PVSKMRMATPLMQA 21
|||||

RESULT 2
Q8SNAO HUMAN PRELIMINARY; PRT; 160 AA.
AC Q8SNAO;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DE 07-FEB-2006, entry version 15.
DE CD74 antigen, isoform c.
GN Name=CD74;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX TISSUE=Primary B-Cells;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.J., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
```

```

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schmerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX TISSUE=Primary B-Cells;
RG NIH MGC Project;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BC024272; AAH24272.1; -; mRNA.
DR HSSP; P04233; 1IIE.
DR Ensembl; ENSG00000019582; Homo sapiens.
SQ SEQUENCE 160 AA; 18328 MW; 95CB3D7226497DE8 CRC64;

Query Match 100.0%; Score 73; DB 2; Length 160;
Best Local Similarity 100.0%; Pred. No. 1.5e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLMQA 15
Db 103 PVSKMRMATPLMQA 117
|||||

RESULT 3
Q5U0J8 HUMAN PRELIMINARY; PRT; 232 AA.
AC Q5U0J8;
DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
DE 07-FEB-2006, entry version 3.
DE CD74 antigen (Invariant polypeptide of major histocompatibility
DE complex, class II antigen-associated).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kalnina N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length CDSs in BD Creator(TM) System Donor
RT vector.";
RL Submitted (OCT-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; BT019505; AAV38312.1; -; mRNA.
DR SMR; Q5U0J8; 134-208.
DR Ensembl; ENSG00000019582; Homo sapiens.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR011988; MHCII_invariant.
SQ SEQUENCE 232 AA; 26399 MW; 5144439D0FD27C99 CRC64;

Query Match 100.0%; Score 73; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVSKMRMATPLMQA 15
Db 103 PVSKMRMATPLMQA 117
|||||

RESULT 4
Q5RFJ4_PONPY
```

```

ID QSRFJ4_PONPY PRELIMINARY; PRT; 232 AA.
AC QSRFJ4;
DT 21-DEC-2004, integrated into UniProtKB/TrEMBL.
DT 21-DEC-2004, sequence version 1.
DT 07-FEB-2006, entry version 5.
DE Hypothetical protein DRFP2p469K1522.
GN Name=DKFP2p469K1522;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Pongo.
OX NCBI_TaxID=9600;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RC The German cDNA Consortium;
RG Bahr A., Lauber J., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA Han M., Wiemann S.;
RA Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC ENBL; CR857162; CAH89463.1; -, mRNA.
DR SMR; QSRFJ4; 134-208.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR011988; MHCII_invariant.
KW Hypothetical protein.
SQ SEQUENCE 232 AA; 26411 MW; D85D7AC72AF7439F CRC64;

Query Match 100.0%; Score 73; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PVSQKRMATPLLMQA 15
| | | | | | | | | | | | | | |
Db 103 PVSQKRMATPLLMQA 117

RESULT 5
HG2A HUMAN
ID _HG2A HUMAN STANDARD; PRT; 296 AA.
AC P04233; Q14597; Q29832; Q8WLPE;
DT 20-MAR-1987, integrated into UniProtKB/Swiss-Prot.
DT 16-APR-2002, sequence version 3.
DT 07-MAR-2006, entry version 74.
DE HLA class II histocompatibility antigen, gamma chain (HLA-DR antigens-
DE associated invariant chain) (Ia antigen-associated invariant chain)
DE (I1) (p33) (CD74 antigen).
GN Name=CD74; Synonyms=DHLA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP MEDLINE=84207945; PubMed=6586420;
RA Strubin M., Mach B., Long B.O.;
RT "The complete sequence of the mRNA for the HLA-DR-associated invariant
RT chain reveals a polypeptide with an unusual transmembrane polarity.";
RL EMBO J. 3:869-872(1984).
RN [2]
RC NUCLEOTIDE SEQUENCE.
RP MEDLINE=86093681; PubMed=3001652;
RA Kudo J., Chao L.-Y., Narni F., Saunders G.F.;
RT "Structure of the human gene encoding the invariant gamma-chain of
RT class II histocompatibility antigens.";
RL Nucleic Acids Res. 13:8827-8841(1985).
RN [3]
RC NUCLEOTIDE SEQUENCE (ISOFORMS LONG AND SHORT).
RN TISSUE=Liver;

RX MEDLINE=86233451; PubMed=3459184;
RA O'Sullivan D.M., Lathammar D., Wilson M.C., Peterson P.A.,
RA Quaranta V.;
RT "Structure of the human Ia-associated invariant (gamma)-chain gene:
RT identification of 5' sequences shared with major histocompatibility
RT complex class II genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4484-4488(1986).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM SHORT).
RC TISSUE=Tonsil;
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE OF 27-96.
RX MEDLINE=84170234; PubMed=6324166;
RA Claesson L., Larhammar D., Raak L., Peterson P.A.;
RT "cDNA clone for the human invariant gamma chain of class II
RT histocompatibility antigens and its implications for the protein
RT structure.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:7395-7399(1983).
RN [6]
RP PROTEIN SEQUENCE OF 97-120.
RX MEDLINE=93078879; PubMed=1448172; DOI=10.1038/360474a0;
RA Riberty J.M., Newcomb J.R., Surman M.J., Barbosa J.A., Cresswell P.;
RT "HLA-DR molecules from an antigen-processing mutant cell line are
RT associated with invariant chain peptides.";
RL Nature 360:474-477(1992).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS) OF 103-117.
RX MEDLINE=96085023; PubMed=7477400; DOI=10.1038/378457a0;
RA Ghosh P., Amaya M., Mellins E., Wiley D.C.;
RT "The structure of an intermediate in class II MHC maturation: CLIP
RT bound to HLA-DR3.";
RL Nature 378:457-462(1995).
RN [8]
RP STRUCTURE BY NMR OF 134-208.
RX MEDLINE=99059718; PubMed=9843486; DOI=10.1093/emboj/17.23.6812;
RA Jaeanoff A., Wagner G., Wiley D.C.;
RT "Structure of a trimeric domain of the MHC class II-associated
RT chaperonin and targeting protein Ii.";
RL EMBO J. 17:6812-6818(1998).
RN [9]
CC -!- FUNCTION: Plays a critical role in MHC class II antigen processing
CC by stabilizing peptide-free class II alpha/beta heterodimers in a
CC complex soon after their synthesis and directing transport of the
CC complex from the endoplasmic reticulum to compartments where
CC peptide loading of class II takes place.
CC -!- SUBUNIT: Nonamer composed of three alpha/beta/gamma heterotrimers.
CC -!- SUBCELLULAR LOCATION: Membrane; single-pass type II membrane
CC protein (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P04233-1; Sequence=Displayed;
CC Name=Short;

```





Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
{2}  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Eye;  
RC  
RA Strausberg R.;  
RL Submitted (MAY-2002) to the ENBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC ENBL; BC030458; AAH30458.1; -; mRNA.  
DR HSPF; P04233; 11IE.  
DR SMR; Q8K058; 118-192.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPRO11988; MHCII\_invariant.  
KW Hypothetical protein.  
SQ SEQUENCE 208 AA; 23590 MW; FF36PF7835228A596 CRC64;

Query Match 84.9%; Score 62; DB 2; Length 208;  
Best Local Similarity 73.3%; Pred. No. 0.0027;  
Matches 11; Conservative 4; Mismatches 0; Indels

Qy 1 PVSKMRMATPLLMQA 15  
| : | | | | : | | |  
Db 87 PLSKMRVATPMMMQA 101

RESULT 9

Q6GT22\_MOUSE PRELIMINARY; PRT; 215 AA.

ID Q6GT22 MOUSE

AC Q6GT22;

DT 10-MAY-2005, integrated into UniProtKB/TrEMBL.

DT 10-MAY-2005, sequence version 1.

DT 07-FEB-2006, entry version 11.

DE Ia-associated invariant chain (Adult male kidney cDNA, RIKEN full-length enriched library, clone:0610006B16 product:Ia-associated invariant chain, full insert sequence) (Cd74 protein).

GN Name=Cd74; Synonyms=Ii;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.

NCBI\_TaxID=10090;

[1]

NCLEOTIDE SEQUENCE:

RP STRAIN=FVB/N; TISSUE=Colon, and Salivary gland;

RC MEDLINE=23288257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Hsieh F., Diatchenko L., Marushin K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywicki M.I., Skalski U., Smalhus D.E., Schnerker A., Schein J.E., Jones S.J.M., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[2]

NCLEOTIDE SEQUENCE.

RP STRAIN=FVB/N; TISSUE=Salivary gland;

RG NIH MGC Project;

RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

RA NUCLEOTIDE SEQUENCE.  
RP STRAIN=C57BL/6J; TISSUE=Kidney;  
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;  
RX Carninci P., Hayashizaki Y.;  
RA "high-efficiency full-length cDNA cloning";  
RL Methods Enzymol. 303:19-44(1999).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX PubMed=16141072; DOI=10.1126/science.1112014;  
RA Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K., Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Fujiki S., Gariboldi M., Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E., Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N., Hill D., Hummel L., Iacono K., Ikeo K., Iwama A., Ishikawa T., Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H., Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K., Kurochkin I.V., Laveau L.F., Lazarevic D., Lipovich L., Liu J., Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L., Matsuda H., Matsuzawa S., Miki H., Mignone P., Miyake S., Morris K., Mottagui-Tabar S., Mulder N., Nakano N., Nakachi H., Ng P., Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O., Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavese G., Pesole G., Petroski N., Piazza S., Reid J., Reid J.F., Ring B.Z., Ringwald M., Rost B., Ruan Y., Salzberg S.L., Sadelain A., Schneider C., Schonbach C., Sekiguchi K., Sempile C.A., Seno S., Sessa L., Sheng Y., Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B., Sperling S., Stupka E., Sugita K., Sultana R., Takenaka Y., Taki K., Tamajima K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A., Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K., Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C., Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J., Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y., Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T., Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N., Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N., Nishio T., Okada M., Plessey C., Shibata K., Shiraki T., Suzuki S., Tagami M., Waki K., Watanabe A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y.;  
RL "The transcriptional landscape of the mammalian genome";  
RN Science 309:1559-1563(2005).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX PubMed=16141073; DOI=10.1126/science.1112009;  
RG RIKEN Genome Exploration Research Group, and Genome Science Group (Genome Network Core Team) and the FANTOM Consortium;  
RL "Antisense Transcription in the Mammalian Transcriptome";  
RN Science 309:1564-1566(2005).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Gasterland T., Gariboldi M., Gissi C., Forrest A., Frazer K.S., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Negashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reid J., Reid J.F., Ring B.Z., Ringwald M., Sadelain A., Schneider C., Sempile C.A., Sempile M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
RL [7]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito K., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldi M., Brownstein M.J., Bult C., Fletcher C., Fujita M., Kamiya M., Lee N.H., Gustincich S., Hill D., Hofmann M., Hume D.A., Kiyosawa H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombert P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S., Hayashizaki Y.;  
RL "Functional annotation of a full-length mouse cDNA collection";  
RN Nature 409:685-690(2001).  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
RL "Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes";  
RN Genome Res. 10:1617-1630(2000).  
RN [9]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujitake S., Inoue K., Togawa Y., Izawa M., Ohata E., Watanabe M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsunaga S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
RL "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer";  
RN Genome Res. 10:1757-1771(2000).  
RN [10]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=C57BL/6J; TISSUE=Kidney;  
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H., Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M., Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F., Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H., Kawai J., Kojima Y., Komno H., Kouda M., Koya S., Kurihara C.,

RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saiko H., Saiko R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN [11]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RG NIH MGC Project;

Query Match 83.6%; Score 61; DB 2; Length 215;  
 Best Local Similarity 85.7%; Pred. No. 0.0044;  
 Matches 12; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PVSKMRMATPLLMQ 14  
 |||:|||||:  
 Db 86 PVSQMRMATPLLMR 99  
 |||:|||||:

RESULT 10  
 HG2A MOUSE  
 ID HG2A MOUSE STANDARD; PRT; 279 AA.  
 AC P04441; O19452;  
 DT 13-AUG-1987, integrated into UniProtKE/Swiss-Prot.  
 DT 01-NOV-1991, sequence version 3.  
 DT 07-MAR-2006, entry version 63.  
 DE H-2 class II histocompatibility antigen, gamma chain (MHC class II-  
 DE associated invariant chain) (Ia antigen-associated invariant chain)  
 DE (I1) (CD74 antigen).  
 GN Name=CD74; Synonyms=I1;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
 OC Muridae; Muridae; Murinae; Mus.  
 OX NCBI TaxID=10090;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AKR; TISSUE=Liver;  
 RX MEDLINE=89098406; PubMed=2492095;  
 RA Zhu L., Jones P.P.;  
 RT "Complete sequence of the murine invariant chain (I1) gene.";  
 RL Nucleic Acids Res. 17:447-448(1989).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE, AND ALTERNATIVE SPLICING.  
 RC STRAIN=AKR;  
 RX MEDLINE=87275961; PubMed=3038530;  
 RA Koch N., Lauer W., Habicht J., Dobberestein B.;  
 RT "Primary structure of the gene for the murine Ia antigen-associated  
 RT invariant chains (I1). An alternatively spliced exon encodes a  
 RT cysteine-rich domain highly homologous to a repetitive sequence of  
 RT thyroglobulin.";  
 RL EMBO J. 6:1677-1683(1987).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] (ISOFORM LONG).  
 RC STRAIN=129; TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.D., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE OF 1-58.  
 RA Stone J., Perry R., Todd J.A., McDevitt H.O.;  
 RT "Nucleotide sequences of the murine Ia-associated invariant chain (I1)  
 RT and I-E (H-2S, Beta) chain expressible cDNA clones.";  
 RL Submitted (MAR-1988) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE OF 47-191 AND 256-279.  
 RX MEDLINE=84207946; PubMed=6327293;  
 RA Singer P.A., Lauer W., Dembic Z., Mayer W.E., Lipp J., Koch N.,  
 RA Hammerling G., Klein J., Dobberestein B.;  
 RT "Structure of the murine Ia-associated invariant (I1) chain as deduced  
 RT from a cDNA clone.";  
 RL EMBO J. 3:873-877(1984).  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE OF 1-25.  
 RX MEDLINE=90257363; PubMed=2111346;  
 RA Eades A.-M., Litfin M., Rahmsdorf H.J.;  
 RT "The IFN-gamma response of the murine invariant chain gene is mediated  
 RT by a complex enhancer that includes several MHC class II consensus  
 RT elements.";  
 RL J. Immunol. 144:4399-4409(1990).  
 RN [7]  
 RP MUTAGENESIS OF SER-265.  
 RX MEDLINE=8814436; PubMed=3422739;  
 RA Miller J., Hatch J.A., Simonis S., Cullen S.E.;  
 RT "Identification of the glycosaminoglycan attachment site of mouse  
 RT invariant-chain proteoglycan core protein by site-directed  
 RT mutagenesis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1359-1363(1988).  
 RN [8]  
 RP FUNCTION: Plays a critical role in MHC class II antigen processing  
 CC by stabilizing peptide-free class II alpha/beta heterodimers in a  
 CC complex soon after their synthesis and directing transport of the  
 CC complex from the endoplasmic reticulum to compartments where  
 CC peptide loading of class II takes place.  
 CC SUBUNIT: Nonamer composed of three alpha/beta/gamma heterotrimers  
 CC (By similarity).  
 CC SUBCELLULAR LOCATION: Membrane; single-pass type II membrane  
 CC protein (Potential).  
 CC ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC ISOID=P04441-1; Sequence=Displayed;  
 CC Name=Short;  
 CC ISOID=P04441-2; Sequence=VSP 005332;  
 CC SIMILARITY: Contains 1 thyroglobulin type-1 domain.  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License

EMBL: X00496; CAA25191.1; -; mRNA.  
 EMBL: X07129; CAA30141.1; -; mRNA.  
 EMBL: X05428; CAA29010.1; -; Genomic DNA.  
 EMBL: X05429; CAA29012.1; -; Genomic DNA.  
 EMBL: X05430; CAA37297.1; -; Genomic DNA.  
 EMBL: BC003476; AA03476.1; -; mRNA.  
 EMBL: X13414; -; NOT ANNOTATED CDS; Genomic DNA.  
 EMBL: M35872; AAA37897.1; -; Genomic DNA.  
 PIR: B27866; HLHMSG.  
 HSSP: P04233; 1ICF.  
 SMR: P04441; 117-191, 193-256.  
 Ensembl: ENSMUSG00000024610; Mus musculus.  
 MGI: MG1:96534; I1.  
 GO: GO:0005783; C:Endoplasmic reticulum; IDA.  
 GO: GO:0009897; C:external side of plasma membrane; IDA.  
 GO: GO:0005794; C:Golgi apparatus; IDA.  
 GO: GO:0005887; C:integral to plasma membrane; TAS.  
 GO: GO:0005764; C:lysosome; IDA.







```

DR EMBL; BC059152; AAH59152.1; -; mRNA.
DR SMR; Q6G770; 118-192.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR011988; MHCII_invariant.
SQ SEQUENCE 216 AA; 24503 MW; 33677FA259B84F8E CRC64;

Query Match 80.8%; Score 59; DB 2; Length 216;
Best Local Similarity 85.7%; Pred. No. 0.011;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKMRMATPLLMQ 14
   ||| ||||| |||
Db 87 PVSPMRMATPLLMR 100

RESULT 15
HG2A RAT STANDARD; PRT; 280 AA.
AC P10247;
DT 01-JUL-1989, integrated into UniProtKB/Swiss-Prot.
DT 01-FEB-1991, sequence version 2.
DT 07-MAR-2006, entry version 53.
DE H-2 class II histocompatibility antigen, gamma chain (MHC class II-
DE associated invariant chain) (Ia antigen-associated invariant chain)
DE (II) (CD74 antigen).
GN Name=Cd74;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA]
RC STRAIN=PVG X DA; TISSUE=Spleen;
RX MEDLINE=89282409; PubMed=2499873;
RA McKnight A.J., Mason D.W., Barclay A.N.;
RT "Sequence of a rat MHC class II-associated invariant chain cDNA clone
RL containing a 64 amino acid thyroglobulin-like domain.";
RL Nucleic Acids Res. 17:3983-3984(1989).
[2]
RP NUCLEOTIDE SEQUENCE [MRNA] OF 1-192 AND 257-280.
RC STRAIN=Lewis;
RX MEDLINE=89098337; PubMed=3264906;
RA Henkes W., Syha J., Reske K.;
RT "Nucleotide sequence of rat invariant gamma chain cDNA clone pLR gamma
RT 34.3.";
RL Nucleic Acids Res. 16:11822-11822(1988).
CC -!- FUNCTION: Plays a critical role in MHC class II antigen processing
CC by stabilizing peptide-free class II alpha/beta heterodimers in a
CC complex soon after their synthesis and directing transport of the
CC complex from the endoplasmic reticulum to compartments where
CC peptide loading of class II takes place.
CC -!- SUBUNIT: Nonamer composed of three alpha/beta/gamma heterotrimers
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Membrane; single-pass type II membrane
CC protein (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P10247-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P10247-2; Sequence=VSP 005333;
CC -!- SIMILARITY: Contains 1 thyroglobulin type-1 domain.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; X14254; CAA32468.1; -; mRNA.
CC EMBL; X13044; CAA31450.1; -; mRNA.
CC PIR; S04362; S04362.
CC HSSP; P04233; 1ICF.
CC SMR; P10247; 118-192, 194-257.

```

---

```

DR Ensembl; ENSRNOG00000018735; Rattus norvegicus.
DR RGD; 2313; Cd74.
DR GO; GO:0019955; P:Cytokine binding; ISS.
DR GO; GO:0008283; P:cell proliferation; ISS.
DR GO; GO:0043066; P:negative regulation of apoptosis; ISS.
DR GO; GO:0001516; P:prostaglandin biosynthesis; ISS.
DR GO; GO:0043030; P:regulation of macrophage activation; ISS.
DR InterPro; IPR011988; MHCII_invariant.
DR InterPro; IPR000716; Thyroglobulin_1.
DR Pfam; PF00086; Thyroglobulin_1; 1.
DR SMART; SM00211; TV; 1.
DR PROSITE; PS00484; THYROGLOBULIN_1_1; 1.
DR PROSITE; PS51162; THYROGLOBULIN_1_2; 1.
KW Alternative splicing; Chaperone; Glycoprotein; Immune response;
KW Membrane; Proteoglycan; Signal-anchor; Transmembrane.
FT CHAIN 1 280
FT H-2 class II histocompatibility antigen,
FT gamma chain.
FT /FTID=PRO 0000067956.
FT TOPO_DOM 1 30 Cytoplasmic (Potential).
FT TRANSMEM 31 56 Signal-anchor for type II membrane
FT protein (Potential).
FT TOPO_DOM 57 280 Extracellular (Potential).
FT DOMAIN 194 255 Thyroglobulin type-1.
FT CARBOHYD 114 114 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 120 120 N-linked (Xyl...) (Potential).
FT CARBOHYD 266 266 O-linked (Xyl...) (glycosaminoglycan) (By
FT similarity).
FT DISULFID 197 216 By similarity.
FT DISULFID 227 234 By similarity.
FT DISULFID 236 255 By similarity.
FT VARSPLIC 193 256 Missing (in isoform Short).
FT /FTID=VSP 005333.
SQ SEQUENCE 280 AA; 31642 MW; D935D169A98B5732 CRC64;

Query Match 80.8%; Score 59; DB 1; Length 280;
Best Local Similarity 85.7%; Pred. No. 0.014;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKMRMATPLLMQ 14
   ||| ||||| |||
Db 87 PVSPMRMATPLLMR 100

RESULT 16
Q764N1 PIG
ID Q764N1_PIG PRELIMINARY; PRT; 214 AA.
AC Q764N1;
DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.
DT 05-JUL-2004, sequence version 1.
DT 07-FEB-2006, entry version 9.
DE CD74 antigen.
GN Name=Cd74;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=14681463; DOI=10.1093/nar/gkh037;
RA Uenishi H., Eguchi T., Suzuki K., Sawazaki T., Toki D., Shinkai H.,
RA Okumura N., Hamasima N., Awata T.;
RA "PEDE (Pig EST Data Explorer): construction of a database for ESTs
RT derived from porcine full-length cDNA libraries.";
RL Nucleic Acids Res. 32:D484-D488(2004).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL; AB116558; BAD06312.1; -; mRNA.
CC GO; GO:0016020; C:membrane; IEA.
CC GO; GO:0006955; P:immune response; IEA.
CC InterPro; IPR011988; MHCII_invariant.

```



SQ SEQUENCE 214 AA; 24118 MW; E5347DE04885524F CRC64;

Query Match 76.7%; Score 56; DB 2; Length 214;  
Best Local Similarity 66.7%; Pred. No. 0.04;  
Matches 10; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKRMATPLLMQA 15  
|:||||:|:||||  
DB 87 PLSKMRVSAFMLMQA 101

RESULT 17

Q9SJV5\_ARATH PRELIMINARY; PRT; 340 AA.

AC Q9SJV5; Q9XFI4;  
DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.  
DT 01-JUN-2002, sequence version 2.  
DT 07-FEB-2006, entry version 22.  
DE TATC (CpTratC).  
GN Name=apg2; OrderedLocusNames=At2g01110;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Town C.D., Haas B.J., Wu D., Maiti R., Hannick L.I., Chan A.P.,  
RA Tallon L.J., Rooney T., Uterback T.R., VanAken S.E., Feldblyum T.V.,  
RA White O., Fraser C.M.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Lin X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RA Town C.D., Kaul S.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21417794; PubMed=11526245; DOI=10.1073/pnas.181304598;  
RA Mochizuki R., Nagata N., Ito T., Takahashi S., Hobo T., Yoshida S.,  
RA Shinozaki K.;  
RT "An essential role of a TatC homologue of a Delta pH- dependent  
protein transporter in thylakoid membrane formation during chloroplast  
development in Arabidopsis thaliana.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:10499-10504 (2001).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,  
RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,  
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,  
RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22088475; PubMed=12093376;  
RA Haas B.J., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;  
RT "Full-length messenger RNA sequences greatly improve genome  
annotation.";  
RL Genome Biol. 3:RESEARCH0029-RESEARCH0029 (2002).  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [8]  
RP NUCLEOTIDE SEQUENCE.  
RA Summer E.J., McCaffery M.W., Cline K.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [9]

RP NUCLEOTIDE SEQUENCE.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,  
RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,  
RA Kaniya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,  
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,  
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; AF360172; AAK25882.1; -; mRNA.  
EMBL; AC006837; AAF18659.2; -; Genomic\_DNA.  
EMBL; AB054096; BAB62074.1; -; mRNA.  
EMBL; AF145045; AAD33946.1; -; mRNA.  
EMBL; AY084784; AAM61351.1; -; mRNA.  
EMBL; AY056347; AAL07196.1; -; mRNA.  
PIR; T44583; T44583.  
TAIR; At2g01110; -;  
DR InterPro; IPR002033; Translocase.  
DR Pfam; PF00902; TatC; 1.  
DR PRINTS; PR01840; TATCFAMILY.  
DR TIGRFAMs; TIGR00945; tatC; 1.  
DR PROSITE; PS01218; TATC; 1.  
SQ SEQUENCE 340 AA; 37367 MW; 1429C265BFEDB281 CRC64;

Query Match 57.5%; Score 42; DB 2; Length 340;  
Best Local Similarity 66.7%; Pred. No. 34;  
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKRMATPLLM 12  
|:|:|:|:|:|:|  
DB 314 PVTQMLLATPLL 325

RESULT 18

Q3GAU7\_9DELTA PRELIMINARY; PRT; 1054 AA.

AC Q3GAU7;  
DT 08-NOV-2005, integrated into UniProtKB/TrEMBL.  
DT 08-NOV-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Hydrophobe/amphiphile efflux-1 HAE1.  
GN ORFNames=ProDRAFT\_2759;  
OS Pelobacter propionicus DSM 2379;  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfuromonadales;  
OC Pelobacteraceae; Pelobacter.  
OX NCBI\_TaxID=338966;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=DSM 2379;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Dettler J.C., Glavina T.,  
RA Hammon N., Israni S., Pitluck S., Richardson P.;  
RT "Sequencing of the draft genome and assembly of Pelobacter propionicus  
DSM 2379.";  
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=DSM 2379;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RT "Annotation of the draft genome assembly of Pelobacter propionicus DSM  
2379.";  
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License

```
CC EMBL; AAJH01000006; EA036965.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transporter activity; IEA.
DR InterPro; IPR001036; Acrlflin_res.
DR InterPro; IPR004764; HAE1.
DR Pfam; PF00873; ACR_cran; 1.
DR PRINTS; PR00702; ACRIFLAVINRP.
DR TIGRFAMs; TIGR00915; 2A0602.1.
SQ SEQUENCE 1054 AA; 114349 MW; 5F57EC6404F4F839 CRC64;

Query Match 57.5%; Score 42; DB 2; Length 1054;
Best Local Similarity 61.5%; Pred. No. 1.le-02;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 3 SKRMATPLLMQA 15
Db 109 NKQLATPLLPQA 121

RESULT 19
Q3AZT7 SYN99 PRELIMINARY; PRT; 110 AA.
AC Q3AZT7;
DT 22-NOV-2005, integrated into UniProtKB/TrEMBL.
DT 22-NOV-2005, sequence version 1.
DT 07-FEB-2006, entry version 3.
DE YGGT family, conserved hypothetical integral membrane protein.
GN ORFNames=Syncc9902_0420;
OS Synchococcus sp. (strain CC9902).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=316279;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CC9902;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Iserani S., Pitluck S., Martinez M., Schmutz J., Larimer F.,
RA Land M., Kyripides N., Ivanova N., Richardson P.,
RA "Complete sequence of Synchococcus sp. CC9902";
RL Submitted (AUG-2005) to the EMBL/GenBank/DBJ databases.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC EMBL; CP000097; AB25390.1; -; Genomic DNA.
DR GO; GO:0016020; C:membrane; IEA.
KW Hypothetical protein.
SQ SEQUENCE 110 AA; 12043 MW; A220BA2C0EA9653F CRC64;

Query Match 56.2%; Score 41; DB 2; Length 110;
Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 5 MRMATPLLMQA 15
Db 1 MRFVTPLLLQA 11

RESULT 20
Y291_PASMU STANDARD; PRT; 238 AA.
AC Q9CNY0;
DT 22-AUG-2003, integrated into UniProtKB/Swiss-Prot.
DT 01-JUN-2001, sequence version 1.
DT 07-MAR-2006, entry version 14.
DE Hypothetical protein PM0291 precursor.
GN OrderedLocusNames=PM0291;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_TaxID=747;
```

```
[1]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC EMBL; AE006064; AAK02375.1; -; Genomic DNA.
DR GenomeReviews; AE004439.GR; PM0291.
DR BioCyc; PMUL747:PM0291-MONOMER; -.
KW Complete proteome; Hypothetical protein; Signal.
FT SIGNAL 1 22 potential.
FT CHAIN 23 238 Hypothetical protein PM0291.
FT /FTID=PRO_0000014174.
SQ SEQUENCE 238 AA; 26984 MW; 9F507B621AA5B1AE CRC64;

Query Match 56.2%; Score 41; DB 1; Length 238;
Best Local Similarity 57.1%; Pred. No. 36;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PVSQKRMATPLLMQ 14
Db 218 PESQOKLATPLLNQ 231

RESULT 21
Q5YUF1 NOCFA PRELIMINARY; PRT; 471 AA.
AC Q5YUF1;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 21-FEB-2006, entry version 15.
DE Putative transporter.
GN OrderedLocusNames=nfa33430;
OS Nocardia farcinica.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Nocardiaceae; Nocardia.
OX NCBI_TaxID=37329;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IFM 10152;
RX PubMed=15466710; DOI=10.1073/pnas.0406410101;
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,
RA Shiba T., Hattori M.;
RT "The complete genomic sequence of Nocardia farcinica IFM 10152.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930(2004).
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (By
CC similarity).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC EMBL; AP006618; BAD58190.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0015520; F:tetracycline:hydrogen antiporter activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0015904; P:tetracycline transport; IEA.
DR GO; GO:0008810; P:transport; IEA.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR011701; MFS_1.
DR InterPro; IPR001411; TCR_TetB.
DR Pfam; PF07690; MFS_1; 1.
DR PRINTS; PR01036; TCRTEB.
DR PROSITE; PS50850; MFS; 1.
KW Complete proteome; Membrane; Transmembrane; Transport.
SQ SEQUENCE 471 AA; 47603 MW; 7063445CE608F0EA CRC64;

Query Match 56.2%; Score 41; DB 2; Length 471;
```

Best Local Similarity 40.8%; Pred. No. 75;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 PVSKRMATPLMQA 15  
Db 269 PVGAQLVTPWIMQS 283

## RESULT 22

PHAF\_RHOCA STANDARD; PRT; 92 AA.  
AC O68038;  
DT 30-MAY-2000, integrated into UniProtKB/Swiss-Prot.  
DT 01-AUG-1998, sequence version 1.  
DT 07-FEB-2006, entry version 26.  
DE Probable K(+)/H(+) antiporter subunit F (pH adaptation potassium efflux system protein F) (Pha system subunit F).  
GN Namesphaf;  
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Rhodobacter.  
OX NCBI\_TaxID=1061;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RC STRAIN=SB1003 / St Louis;  
RX MEDLINE=97404404; PubMed=9256491;  
RA Vlcek C., Paces V., Maltsev N., Paces J., Haselkorn R., Ponstein M.;  
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388 (1997).  
CC -!- FUNCTION: Part of a K(+) efflux system involved in pH adaptation (by similarity).  
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (Potential).  
CC -!- SIMILARITY: Belongs to the CPA3 antiporters (TC 2.A.63) subunit F family.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: AF010496; AAC16124.1; -; Genomic\_DNA.  
DR PR; T03471; T03471.  
DR InterPro; IPR007208; MrpF\_Phaf.  
DR Pfam; PF04066; MrpF\_Phaf; 1.  
KW Antiport; Hydrogen ion transport; Ion transport; Membrane; Potassium; Potassium transport; Transmembrane; Transport.  
FT CHAIN 1 92 Probable K(+)/H(+) antiporter subunit F.  
FT /FTID=PRO\_0000087745.  
FT Potential.  
FT TRANSMEM 4 24 Potential.  
FT TRANSMEM 37 57 Potential.  
FT TRANSMEM 62 82 Potential.  
SQ SEQUENCE 92 AA; 9790 MW; 953B3509FB542A2A CRC64;

Query Match 54.8%; Score 40; DB 1; Length 92;  
Best Local Similarity 57.1%; Pred. No. 20;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 VSKRMATPLMQA 15  
Db 53 VIGRMGTFFLFEA 66

## RESULT 23

Q3QS54\_9RHOB PRELIMINARY; PRT; 335 AA.  
AC Q3QS54;  
DT 25-OCT-2005, integrated into UniProtKB/TrEMBL.  
DT 25-OCT-2005, sequence version 1.  
DT 07-FEB-2006, entry version 3.  
DE Inner-membrane translocator precursor.  
GN ORFNames=RoseDRAFT\_2168;  
OS Silicibacter sp. TM1040.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;

OC Rhodobacteraceae; Silicibacter.  
OX NCBI\_TaxID=292414;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TM1040;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T., Hamon N., Israni S., Pitluck S., Richardson P.;  
RT "Sequencing of the draft genome and assembly of Silicibacter sp. TM1040."  
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TM1040;  
RG US DOE Joint Genome Institute (JGI-ORNL);  
RA Larimer F., Land M.;  
RT "Annotation of the draft genome assembly of Silicibacter sp. TM1040."  
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TM1040;  
RG US DOE Joint Genome Institute (JGI-PGF);  
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T., Hamon N., Israni S., Pitluck S., Richardson P.;  
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.  
CC  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC  
CC EMBL: AAFG02000005; EANS56679.1; -; Genomic\_DNA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005215; F:transporter activity; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR001851; Bac\_inmem\_transp.  
DR Pfam; PF02653; BPD\_transp\_2; 1.  
KW Signal.  
FT SIGNAL 1 56 Potential.  
SQ SEQUENCE 335 AA; 35328 MW; 3DD256F4FAB77A60 CRC64;  
Query Match 54.8%; Score 40; DB 2; Length 335;  
Best Local Similarity 72.7%; Pred. No. 81;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 SKRMATPLLM 13  
Db 12 STVRMATPLLL 22

RESULT 24  
Q7Q9V4\_ANOGA PRELIMINARY; PRT; 350 AA.  
AC Q7Q9V4;  
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.  
DT 07-DEC-2004, sequence version 2.  
DT 07-FEB-2006, entry version 12.  
DE ENSANGP0000021807 (Fragment).  
GN ORFNames=ENSANG0000001318;  
OS Anopheles gambiae str. PEST.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;  
OC Anophelinae; Anopheles.  
OX NCBI\_TaxID=180454;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PEST;  
RG The Anopheles gambiae Sequence Committee;  
RT "Anopheles gambiae re-annotation."  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP NUCLEOTIDE SEQUENCE.

```

RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AAB01008900; EAA09481.2; -; Genomic_DNA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR002048; EF_hand_Ca_Bd.
DR InterPro; IPR004181; ZnF_MiZ.
DR Pfam; PF02891; zf-MiZ; 1.
DR PROSITE; PS00018; EF_HAND_1; UNKNOWN_1.
DR PROSITE; PS51044; ZF_SP_RING; 1.
FT NON_TER 1
FT NON_TER 350
SQ SEQUENCE 350 AA; 38908 MW; 3DF953760C0688B8 CRC64;

Query Match 54.8%; Score 40; DB 2; Length 350;
Best Local Similarity 70.0%; Pred. No. 85;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PVSQWRMATP 10
Db 208 PLGKWRMTTP 217
|:|||||

RESULT 25
O22188 ARATH PRELIMINARY; PRT; 543 AA.
ID O22188 ARATH PRELIMINARY; PRT; 543 AA.
AC O22188;
DT 01-JAN-1998, integrated into UniProtKB/TrEMBL.
DT 01-JAN-1998, sequence version 1.
DT 21-FEB-2006, entry version 31.
DE Putative cytochrome P450.
GN OrderedLocusNames:At2g23190;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AC002391; AAB87112.1; -; Genomic_DNA.
DR PIR; T00513; T00513.
DR HSSP; P14779; LUPZ.
DR TAIR; At2g23190; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0020037; F:heme binding; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0004497; F:monooxygenase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR001128; Cytochrome_P450.
DR InterPro; IPR002401; EP450I.
DR PANTHER; PTHR19383; Cytochrome_P450; 1.

```

```

DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00463; EP450I.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME P450; UNKNOWN_1.
KW Heme; Iron; Membrane; Metal-binding; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 543 AA; 62777 MW; 06332D6C090FDB12 CRC64;

Query Match 54.8%; Score 40; DB 2; Length 543;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PVSQWRMATPLL 12
Db 515 PNMWRPATPLL 526
|::|||

RESULT 26
Q61DK9 CAEBR PRELIMINARY; PRT; 675 AA.
ID Q61DK9 CAEBR PRELIMINARY; PRT; 675 AA.
AC Q61DK9;
DT 23-NOV-2004, integrated into UniProtKB/TrEMBL.
DT 23-NOV-2004, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein CBG12444.
GN Name=CBG12444;
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=AF16;
RX PubMed=14624247; DOI=10.1371/journal.pbio.0000045;
RA Stein L.D., Bao Z., Blasiar D., Blumenthal T., Brent M.R., Chen N.,
RA Chinwalla A., Clarke L., Clee C., Coghlan A., Coulson A.,
RA D'Eustachio P., Fitch D.H.A., Fulton L.A., Fulton R.E.,
RA Griffiths-Jones S., Harris T.W., Hillier L.W., Kamath R.,
RA Kuwabata P.E., Mardis E.R., Marra M.A., Miner T.L., Minx P.,
RA Mullikin J.C., Plumb R.W., Rogers J., Schein J.E., Sohrmann M.,
RA Spieth J., Stajich J.E., Wei C., Willey D., Wilson R.K., Durbin R.,
RA Waterston R.H.;
RA "The genome sequence of Caenorhabditis briggsae: a platform for
RT comparative genomics.";
RL PLoS Biol. 1:166-192(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; CAAC01000060; CAE67038.1; -; Genomic_DNA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR004181; ZnF_MiZ.
DR Pfam; PF02891; zf-MiZ; 1.
DR PROSITE; PS51044; ZF_SP_RING; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 675 AA; 75900 MW; BF2633230A44B877 CRC64;

Query Match 54.8%; Score 40; DB 2; Length 675;
Best Local Similarity 63.6%; Pred. No. 1.7e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PVSQWRMATPLL 11
Db 267 PLSKIRMTTPV 277
|::|||

RESULT 27
UVRD_SALTY
ID UVRD_SALTY STANDARD; PRT; 720 AA.
AC Q05311; Q9L6P3;
DT 01-OCT-1994, integrated into UniProtKB/Swiss-Prot.

```

```
DT 29-AUG-2001, sequence version 2.
DE 07-MAR-2006, entry version 41.
DE DNA helicase II (EC 3.6.1.-).
GN Name=uvrD; OrderedLocusNames=STM3951; ORFNames=STWD1.39;
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=L72 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT L72."
RL J. Biol. Chem. 268.14071-14080 (1993).
CC -i- FUNCTION: Has both ATPase and helicase activities. Unwinds DNA
CC duplexes with 3' to 5' polarity with respect to the bound strand
CC and initiates unwinding most effectively when a single-stranded
CC region is present. Involved in the post-incision events of
CC nucleotide excision repair and methyl-directed mismatch repair.
CC -i- SIMILARITY: Belongs to the helicase family. UvrD subfamily.
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL: AF233324; AAF33441.1; -: Genomic DNA.
CC EMBL: AE008884; AAL22795.1; -: Genomic DNA.
CC EMBL: L11043; AAA02965.1; -: Unassigned DNA.
CC HSSP: P09980; 1UAA.
CC GenomeReviews; AE006468 GR; STM3951.
CC StyGene; SG10414; uvrD.
CC BioCyc; STYP9287:STM3951-MONOMER; -.
DR InterPro; IPR005753; UvrD.
DR InterPro; IPR00212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 1.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR01075; uvrD; 1.
KW ATP-binding; Complete proteome; DNA damage; DNA repair;
KW DNA replication; DNA-binding; Helicase; Hydrolase; Nucleotide-binding;
KW SOS response.
FT CHAIN 1 720 DNA helicase II.
FT FTID=PRO_0000102073.
FT NP_BIND 29 36 ATP (Potential).
FT SEQUENCE 720 AA; 81981 MW; C56D60296392511C CRC64;
Query Match 54.8%; Score 40; DB 1; Length 720;
Best Local Similarity 72.7%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 PVSKRMATPL 11
Db 653 PVSHQRMGTPL 663
RESULT 28
Q57HQ6 SALCH PRELIMINARY; PRT; 720 AA.
AC Q57HQ6;
DC 10-MAY-2005, integrated into UniProtKB/TrEMBL.
DT 10-MAY-2005, sequence version 1.
```

```
DT 07-FEB-2006, entry version 4.
DE DNA-dependent ATPase I and helicase II.
DE Name=uvrD; OrderedLocusNames=SC3850; ORFNames=SCH_3850;
GN Salmonella choleraesuis.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=591;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=SC-B67;
RX PubMed=15781495; DOI=10.1093/nar/gki297;
RA Chiu C.-H., Tang P., Chu C., Hu S., Bao Q., Yu J., Chou Y.-Y.,
RA Wang H.-S., Lee Y.-S.;
RT "The genome sequence of Salmonella enterica serovar Choleraesuis, a
RT highly invasive and resistant zoonotic pathogen."
RL Nucleic Acids Res. 33:1690-1698 (2005).
CC
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC
CC EMBL: AE017220; AAX67756.1; -: Genomic DNA.
CC GO; GO:0005737; C:cytoplasm; IEA.
CC GO; GO:0005524; F:ATP binding; IEA.
CC GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
CC GO; GO:0003677; F:DNA binding; IEA.
CC GO; GO:0004386; F:helicase activity; IEA.
CC GO; GO:0006281; P:DNA repair; IEA.
CC GO; GO:0006268; P:DNA unwinding during replication; IEA.
CC InterPro; IPR005753; UvrD.
CC InterPro; IPR00212; UvrD-helicase.
CC PANTHER; PTHR11070; UvrD-helicase; 1.
CC Pfam; PF00580; UvrD-helicase; 1.
CC TIGRFAMs; TIGR01075; uvrD; 1.
KW Complete proteome; Helicase.
SQ SEQUENCE 720 AA; 81953 MW; CBBE8D126B03D80C CRC64;
Query Match 54.8%; Score 40; DB 2; Length 720;
Best Local Similarity 72.7%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 PVSKRMATPL 11
Db 653 PVSHQRMGTPL 663
RESULT 29
Q5PKM2 SALPA PRELIMINARY; PRT; 720 AA.
AC Q5PKM2;
DC 04-JAN-2005, integrated into UniProtKB/TrEMBL.
DT 04-JAN-2005, sequence version 1.
DT 07-FEB-2006, entry version 8.
DE DNA helicase II.
GN Name=uvrD; OrderedLocusNames=SPA3792;
OS Salmonella paratyphi-a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=54388;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 9150 / SARB42;
RX PubMed=15531882; DOI=10.1038/ng1470;
RA McClelland M., Sanderson K.E., Clifton S.W., Latreille P.,
RA Porwollik S., Sabo A., Meyer R., Bieri T., Ozerky P., McLellan M.,
RA Harkins C.R., Wang C., Nguyen C., Berghoff A., Elliott G.,
RA Kohlberg S., Strong C., Du F., Carter J., Kremizki C., Layman D.,
RA Leonard S., Sun H., Fulton L., Nash W., Miner T., Minx P.,
RA Delehaanty K., Fronick C., Magrini V., Nhan M., Warren W., Florea L.,
RA Spieth J., Wilson R.K.;
RT "Comparison of genome degradation in Paratyphi A and Typhi, human-
RT restricted serovars of Salmonella enterica that cause typhoid."
RL Nat. Genet. 36:1268-1274 (2004).
```

```
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; CP000026; AAV79568.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004386; F:Helicase activity; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006268; P:DNA unwinding during replication; IEA.
DR InterPro; IPR005753; UvrD.
DR InterPro; IPR00212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 1.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR01075; uvrD; 1.
KW Complete proteome; Helicase.
SQ SEQUENCE 720 AA; 81953 MW; CBBE8D126B03D80C CRC64;

Query Match 54.8%; Score 40; DB 2; Length 720;
Best Local Similarity 72.7%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PVSKRMATPL 11
   ||| |||
Db 653 PVSHQRMGTPL 663

RESULT 30
Q823B0_SALTI
ID Q823B0_SALTI PRELIMINARY; PRT; 720 AA.
AC Q823B0; Q7C6W2;
DT 01-MAR-2002, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2002, sequence version 1.
DT 07-MAR-2006, entry version 23.
DE DNA helicase II.
GN Name=uvrD; OrderedLocusNames=STY3608, t3346;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RX DOI=10.1128/JB.185.7.2330-2337.2003;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; AL627278; CAD07941.1; -; Genomic DNA.
DR EMBL; AE014613; AAO70874.1; -; Genomic_DNA.
DR HSPF; P09980; IUNA.
```

```
DR GenomeReviews; AL513382_GR; STY3608.
DR GenomeReviews; AB014613_GR; t3346.
DR BioCyc; SENT209261.T3346-MONOMER-; -.
DR BioCyc; SENT90370.STY3608-MONOMER-; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004003; F:ATP-dependent DNA helicase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004386; F:Helicase activity; IEA.
DR GO; GO:0006281; P:DNA repair; IEA.
DR GO; GO:0006268; P:DNA unwinding during replication; IEA.
DR InterPro; IPR005753; UvrD.
DR InterPro; IPR00212; UvrD-helicase.
DR PANTHER; PTHR11070; UvrD-helicase; 1.
DR Pfam; PF00580; UvrD-helicase; 1.
DR TIGRFAMs; TIGR01075; uvrD; 1.
KW Complete proteome; Helicase.
SQ SEQUENCE 720 AA; 82110 MW; 9E926B873D9CB672 CRC64;

Query Match 54.8%; Score 40; DB 2; Length 720;
Best Local Similarity 72.7%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PVSKRMATPL 11
   ||| |||
Db 653 PVSHQRMGTPL 663

RESULT 31
Q2IRQ8_RHOPA
ID Q2IRQ8_RHOPA PRELIMINARY; PRT; 755 AA.
AC Q2IRQ8;
DT 07-MAR-2006, integrated into UniProtKB/TrEMBL.
DT 07-MAR-2006, sequence version 1.
DT 07-MAR-2006, entry version 1.
DE Aldehyde oxidase and xanthine dehydrogenase, molybdopterin binding.
GN ORFNames=RPB_4416;
OS Rhodopseudomonas palustris HaA2.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Rhodopseudomonas.
OX NCBI_TaxID=316058;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HaA2;
RG US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hammon N., Israni S., Pittluck S., Chain P., Malfatti S., Shin M.,
RA Vergez L., Schmutz J., Larimer F., Land M., Hauser L., Pelletier D.A.,
RA Kyripides N., Anderson I., Oda Y., Harwood C.S., Richardson P.;
RT "Complete sequence of Rhodopseudomonas palustris HaA2.";
RL Submitted (JAN-2006) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
-----
DR EMBL; CP000250; ABD09102.1; -; Genomic DNA.
DR SEQUENCE 755 AA; 79136 MW; 611FD147885CA0F3 CRC64;

Query Match 54.8%; Score 40; DB 2; Length 755;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 PVSKRMATPL 12
   ||:|:|:|
Db 237 PIKVRITPLI 248

RESULT 32
Q9BZT1_HUMAN
ID Q9BZT1_HUMAN PRELIMINARY; PRT; 140 AA.
AC Q9BZT1;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
```

```

DT 07-FEB-2006, entry version 7.
DE PNAS-101.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Yu W.-Q., Sun B.-Z., Chai Y.-B., Zhu F., Liu X.-S., Li Z., Lu F.,
RA Yan W., Yang H., Zhao Z.-L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AF274958; AAK07533.1; -; mRNA.
SQ SEQUENCE 140 AA; 16335 MW; 68CCA7520ED6CE51 CRC64;

Query Match 53.4%; Score 39; DB 2; Length 140;
Best Local Similarity 64.3%; Pred. No. 49;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PVSKMRMATPLLMQ 14
DB 29 PVSFRLATQLLHQ 42

RESULT 33
Q47S32_THEFY PRELIMINARY; PRT; 141 AA.
AC Q47S32;
DT 13-SEP-2005, integrated into UniProtKB/TrEMBL.
DT 13-SEP-2006, entry version 1.
DT 07-FEB-2006, entry version 4.
DE Putative ATP/GTP binding protein.
OS OrderedLocustNames=Tfu_0697;
OS Thermobifida fusca (strain YX).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=269800;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RA US DOE Joint Genome Institute;
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter J.C., Glavina T.,
RA Hannon N., Israeli S., Pitluck S., Di Bartolo G., Chain P., Schmutz J.,
RA Larimer F., Land M., Lykidis A., Richardson P.;
RA "Complete sequence of Thermobifida fusca YX.";
RL Submitted (JUL-2005) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: CP000088; AAZ54735.1; -; Genomic DNA.
DR GO: GO:0006950; P: response to stress; IEA.
KW Complete proteome.
SQ SEQUENCE 141 AA; 14666 MW; CC24A2B46964BA77 CRC64;

Query Match 53.4%; Score 39; DB 2; Length 141;
Best Local Similarity 53.3%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSKMRMATPLLMQ 15
DB 124 PVSXALATPITLDA 138

RESULT 34
Q18767_CAEEI PRELIMINARY; PRT; 304 AA.
AC Q18767;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2005, sequence version 2.
21-FEB-2006, entry version 31.
Hypothetical protein srxx-30.
Name=srxx-30; ORFNames=C51E3.5;
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
CC -!- SUBCELLULAR LOCATION: Membrane; multi-pass membrane protein (By
CC similarity).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: Z78410; CAB01637.2; -; Genomic DNA.
DR PIR: T20133; T20133.
DR Ensembl: C51E3.5; Caenorhabditis elegans.
DR WormBase; WBGene00008251; C51E3.5.
DR WormPep; C51E3.5; C537815.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0001584; F: rhodopsin-like receptor activity; IEA.
DR GO: GO:0007186; P: G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR PROSITE: PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 304 AA; 34320 MW; 28DF98B71CCE198D CRC64;

Query Match 53.4%; Score 39; DB 2; Length 304;
Best Local Similarity 57.1%; Pred. No. 11e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 PVSKMRMATPLLMQ 14
DB 87 PVIQGLTAAPLLMQ 100

RESULT 35
Q8TP38_METAC PRELIMINARY; PRT; 411 AA.
AC Q8TP38;
DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2002, sequence version 1.
DT 07-MAR-2006, entry version 18.
DE Cation efflux system protein.
DR ORFNames=MA_2085;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Ancoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
```

Matches	7; Conservative	2; Mismatches	1; Indels	0; Gaps
QY	1 PVSXMRMATP 10     : :			
Db	378 PVSXKLIRTP 387			
RESULT 37				
Q2U394 ASPOR				
ID	Q2U394 ASPOR PRELIMINARY; PRT; 469 AA.			
AC	Q2U394			
DT	24-JAN-2006, integrated into UniProtKB/TrEMBL.			
DT	24-JAN-2006, sequence version 1.			
DT	07-MAR-2006, entry version 3.			
DE	Coenzyme F420-dependent N5.			
GN	ORFNames=A0090038000132;			
OS	Aspergillus oryzae.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.			
NCBI_TaxID=5062;				
[1]				
RN	NCLEOTIDE SEQUENCE.			
RP	STRAIN=RIB 40;			
RC	PubMed=16372010; DOI=10.1038/nature04300;			
RX	Machida M., Asai K., Sano M., Tanaka T., Kumagai T., Terai G.,			
RA	Kusumoto K., Arima T., Akita O., Kashiwagi Y., Abe K., Gomi K.,			
RA	Horiuchi H., Kitamoto K., Kobayashi T., Takeuchi M., Denning D.W.,			
RA	Galagan J.E., Nierman W.C., Yu J., Archer D.B., Bennett J.W.,			
RA	Bhatnagar D., Cleveland T.E., Fedorova N.D., Gotch O., Horikawa H.,			
RA	Hosoyama A., Ichinomaya M., Igarashi R., Iwashita K., Juvvadi P.R.,			
RA	Kato M., Kato Y., Kin T., Kokubun A., Maeda H., Maeyama N.,			
RA	Maryama J., Nagasaki H., Nakajima T., Oda K., Okada K., Paulsen I.,			
RA	Sakamoto K., Sawano T., Takahashi M., Takase K., Terabayashi Y.,			
RA	Wortman J.R., Yamada O., Yamagata Y., Anazawa H., Hata Y., Koide Y.,			
RA	Komori T., Kovama Y., Minetoki T., Suharnan S., Tanaka A., Isono K.,			
RA	Kunara S., Ogasawara N., Kikuchi H.;			
RT	"Genome sequencing and analysis of Aspergillus oryzae.";			
RL	Nature 438:1157-1161(2005).			
CC	Copyrighted by the UniProt Consortium, see <a href="http://www.uniprot.org/terms">http://www.uniprot.org/terms</a>			
CC	Distributed under the Creative Commons Attribution-NoDerivs License			
CC	EMBL; AP007169; BAE63971.1; -; Genomic_DNA.			
SQ	SEQUENCE 469 AA; 52120 MW; CC1BB0E8A6C9930D CRC64;			
DR				
Query Match	53.4%;	Score 39;	DB 2;	Length 469;
Best Local Similarity	66.7%;	Pred. No. 1.8e+02;		
Matches 10;	Conservative	0;	Mismatches 3;	Indels 2;
Gaps	1;			
QY	1 PVSXMRMATPLLMOA 15 			
Db	223 PASKQR--TPLLFOA 235 			
RESULT 38				
Q5VTA2 HUMAN				
ID	Q5VTA2 HUMAN PRELIMINARY; PRT; 490 AA.			
AC	Q5VTA2;			
DT	07-DEC-2004, integrated into UniProtKB/TrEMBL.			
DT	07-DEC-2004, sequence version 1.			
DT	07-FEB-2006, entry version 11.			
DE	Fas (TNFRSF6) associated factor 1.			
GN	Names=FAP1; ORFNames=RP5-850015.2-003;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
NCBI_TaxID=9606;				
[1]				
RN	NCLEOTIDE SEQUENCE.			
RP	Mashreghi-Mohammadi M.;			
RA	Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.			
RL	[2]			



RP NUCLEOTIDE SEQUENCE.  
RA Moore M.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RA Glithero R.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AL603746; CAH72114.1; -; Genomic DNA.  
DR EMBL; AC091610; CAH72114.1; JOINED; Genomic DNA.  
DR EMBL; AC118557; CAH72114.1; JOINED; Genomic DNA.  
DR EMBL; AL049637; CAH72114.1; JOINED; Genomic DNA.  
DR EMBL; AL359977; CAH72114.1; JOINED; Genomic DNA.  
DR EMBL; AL049637; CA123010.1; -; Genomic DNA.  
DR EMBL; AC091610; CA123010.1; JOINED; Genomic DNA.  
DR EMBL; AC118557; CA123010.1; JOINED; Genomic DNA.  
DR EMBL; AL359977; CA123010.1; JOINED; Genomic DNA.  
DR EMBL; AL603746; CA123010.1; JOINED; Genomic DNA.  
DR EMBL; AL359977; CAH70190.1; -; Genomic DNA.  
DR EMBL; AC091610; CAH70190.1; JOINED; Genomic DNA.  
DR EMBL; AC118557; CAH70190.1; JOINED; Genomic DNA.  
DR EMBL; AL049637; CAH70190.1; JOINED; Genomic DNA.  
DR EMBL; AL603746; CAH70190.1; JOINED; Genomic DNA.  
DR SNR; Q5VTA2; 409-490.  
DR Ensembl; ENSG0000185104; Homo sapiens.  
DR InterPro; IPR006577; UAS.  
DR InterPro; IPR001012; UBX.  
DR Pfam; PF00789; UBX; 1.  
DR SMART; SM00594; UAS; 1.  
DR SMART; SM00166; UBX; 1.  
DR PROSITE; PS00033; UBX; 1.  
SQ SEQUENCE 490 AA; 55998 MW; 99CCT7ED351A0C251 CRC64;  
  
Query Match 53.4%; Score 39; DB 2; Length 490;  
Best Local Similarity 70.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 PVSKRMATP 10  
Db 412 PVSKLRIRTP 421  
|||||:|  
  
RESULT 39  
ID Q9FQY6 CAPAN PRELIMINARY; PRT; 505 AA.  
AC Q9FQY6;  
DT 01-MAR-2001, integrated into UniProtKB/TrEMBL.  
DT 01-MAR-2001, sequence version 1.  
DT 07-FEB-2006, entry version 20.  
DE Cinnamic acid 4-hydroxylase.  
GN Name=C4H;  
OS Capsicum annuum (Bell pepper).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
OC asterids; lamids; Solanales; Solanaceae; Capsicum.  
OX NCBI\_TaxID=4072;  
RN NUCLEOTIDE SEQUENCE.  
RP TISSUE=Root;  
RA Kim K.-W., Lee S.-W.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; AF212318; AAG43824.1; -; mRNA.  
DR GO; GO:0020037; F:heme binding; IEA.  
DR GO; GO:0005506; F:iron ion binding; IEA.  
DR GO; GO:0046872; F:metal ion binding; IEA.

DR GO; GO:0004497; F:monoxygenase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001128; Cytochrome\_P450.  
DR InterPro; IPR002401; EP450I.  
DR PANTHER; PTHR19383; Cytochrome\_P450; 1.  
DR Pfam; PF00067; p450; 1.  
DR PRINTS; PR00463; EP450I.  
DR PRINTS; PR00385; P450.  
DR PROSITE; PS00086; CYTOCHROME\_P450; UNKNOWN 1.  
KW Heme; Iron; Metal-binding; Monoxygenase; Oxidoreductase.  
SQ SEQUENCE 505 AA; 58020 MW; 38E61FCC2E49FE0F CRC64;  
  
Query Match 53.4%; Score 39; DB 2; Length 505;  
Best Local Similarity 70.0%; Pred. No. 2e+02;  
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 4 KMRMATPLLM 13  
Db 366 RLRMATPLLV 375  
:|||||:  
  
RESULT 40  
ID Q6CSN2 YARLI PRELIMINARY; PRT; 524 AA.  
AC Q6CSN2;  
DT 16-AUG-2004, integrated into UniProtKB/TrEMBL.  
DT 16-AUG-2004, sequence version 1.  
DT 07-FEB-2006, entry version 12.  
DE Similar to tr|Q04472 Saccharomyces cerevisiae YMR115w.  
GN OrderedLocusNames=VAL10E16621g;  
OS Yarrowia lipolytica (Candida lipolytica).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Dipodocaceae; Yarrowia.  
OX NCBI\_TaxID=4952;  
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RP STRAIN=CLIB 122 / E 150;  
RX PubMed15229592; DOI=10.1038/nature02579;  
RA Dujon B., Sherman D., Fischer G., Durand P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., V.,  
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe C.,  
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
RA Kerrist A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
RA Swennen D., Tekaija F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpetti C., Gaillardin C., Weissenbach J.,  
RA Wincker P., Souciet J.-L.;  
RT "Genome evolution in yeasts."  
RL Nature 430:35-44 (2004).  
CC -----  
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
CC Distributed under the Creative Commons Attribution-NoDerivs License  
CC -----  
DR EMBL; CR382131; CAG79623.1; -; Genomic DNA.  
KW Complete proteome.  
SQ SEQUENCE 524 AA; 57733 MW; 3BA135B3DB65023A CRC64;  
  
Query Match 53.4%; Score 39; DB 2; Length 524;  
Best Local Similarity 60.0%; Pred. No. 2.1e+02;  
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 PVSKRMATPLMQA 15  
Db 77 PVSKRVFRITLLLA 91  
|||||:|  
  
Search completed: September 9, 2006, 23:00:30  
Job time : 141.468 secs

.



97 40 80.0 892 8 ADU73761 Adw73761 Mouse POS  
98 40 80.0 892 8 ADU68997 Adw68997 House mou  
99 40 80.0 892 9 ADW87412 Adw87412 Mouse POS  
100 40 80.0 892 9 ADZ66429 Adz66429 Mouse POS

## ALIGNMENTS

RESULT 1  
ADI29006  
ID ADI29006 standard; peptide; 12 AA.  
XX AC  
XX ADI29006;  
XX  
XX 15-APR-2004 (first entry)  
XX  
XX Human MHC class II compound spaceholder molecule SEQ ID NO:2.  
XX MHC class II compound; MHC class II component; MHC class II alpha chain;  
KW MHC class II beta chain; peptide binding groove; spaceholder molecule;  
KW effector component; immune response; immune disorder; virucide;  
KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
KW gene therapy; viral infections; bacterial infection; parasitic infection;  
KW neoplastic disease; autoimmunity; toxicity; human.  
XX  
XX Homo sapiens.  
OS Synthetic.  
OS  
XX  
XX WO2004007528-A2.  
XX  
XX 22-JAN-2004.  
XX  
XX 11-JUL-2003; 2003WO-US021767.  
XX  
XX 12-JUL-2002; 2002US-0395494P.  
XX  
XX 22-JUL-2002; 2002US-0397893P.  
XX  
XX (DAND ) DANA FARBER CANCER INST INC.  
XX  
XX Wucherpfenning KW, Seth N;  
PI  
XX  
XX WPI; 2004-122876/12.  
XX  
XX  
XX New MHC class II compound, useful for preparing a composition for  
PT treating immune disorders e.g. viral infections, bacterial infections,  
PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
XX  
XX Claim 11; SEQ ID NO 2; 92pp; English.  
XX  
XX The present invention describes an isolated MHC class II compound (I)  
CC comprising: (a) an MHC class II component comprising at least a portion  
CC of an MHC class II alpha chain and at least a portion of an MHC class II  
CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
CC chain form a peptide binding groove; (b) a spaceholder molecule; and (c)  
CC an effector component, where the effector component is linked to the MHC  
CC class II component. Also described: (1) a pharmaceutical composition  
CC comprising the MHC class II molecule and a carrier; (2) a method of  
CC producing an MHC class II compound; (3) a method of directly identifying  
CC an antigen-specific T cell; (4) a method of regulating an immune response  
CC in a subject; (5) a method of treating an immune disorder in a subject;  
CC (6) a method of regulating an immune response ex vivo in a subject; and  
CC (7) a method of treating an immune disorder ex vivo in a subject. (I) has  
CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive  
CC activities, and can be used in gene therapy. The MHC class II compound  
CC (I) can be used for preparing a composition for treating immune  
CC disorders, e.g., viral infections, bacterial infections, parasitic  
CC infections, neoplastic disease, autoimmunity or toxicity. The present  
CC sequence represents a spaceholder molecule peptide, which can be used in  
XX an MHC class II compound from the present invention.  
XX Sequence 12 AA;  
XX

Query Match 100.0%; Score 50; DB 8; Length 12;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAAAAAAAAA 12  
| | | | | | | | | |  
Db 1 AAAAAAAAAA 12

RESULT 2  
ADI29007  
ID ADI29007 standard; peptide; 13 AA.  
XX  
XX AC ADI29007;  
XX  
XX 15-APR-2004 (first entry)  
XX  
XX Human MHC class II compound spaceholder molecule SEQ ID NO:3.  
XX MHC class II compound; MHC class II component; MHC class II alpha chain;  
KW MHC class II beta chain; peptide binding groove; spaceholder molecule;  
KW effector component; immune response; immune disorder; virucide;  
KW antibacterial; antiparasitic; cytostatic; immunosuppressive;  
KW gene therapy; viral infections; bacterial infection; parasitic infection;  
KW neoplastic disease; autoimmunity; toxicity; human.  
XX  
XX Homo sapiens.  
OS Synthetic.  
OS  
XX  
XX WO2004007528-A2.  
XX  
XX 22-JAN-2004.  
XX  
XX 11-JUL-2003; 2003WO-US021767.  
XX  
XX 12-JUL-2002; 2002US-0395494P.  
XX  
XX 22-JUL-2002; 2002US-0397893P.  
XX  
XX (DAND ) DANA FARBER CANCER INST INC.  
XX  
XX Wucherpfenning KW, Seth N;  
PI  
XX  
XX WPI; 2004-122876/12.  
XX  
XX New MHC class II compound, useful for preparing a composition for  
PT treating immune disorders e.g. viral infections, bacterial infections,  
PT parasitic infections, neoplastic disease, autoimmunity or toxicity.  
XX  
XX Claim 11; SEQ ID NO 3; 92pp; English.  
XX  
XX The present invention describes an isolated MHC class II compound (I)  
CC comprising: (a) an MHC class II component comprising at least a portion  
CC of an MHC class II alpha chain and at least a portion of an MHC class II  
CC beta chain, such that the MHC class II alpha chain and MHC class II beta  
CC chain form a peptide binding groove; (b) a spaceholder molecule; and (c)  
CC an effector component, where the effector component is linked to the MHC  
CC class II component. Also described: (1) a pharmaceutical composition  
CC comprising the MHC class II molecule and a carrier; (2) a method of  
CC producing an MHC class II compound; (3) a method of directly identifying  
CC an antigen-specific T cell; (4) a method of regulating an immune response  
CC in a subject; (5) a method of treating an immune disorder in a subject;  
CC (6) a method of regulating an immune response ex vivo in a subject; and  
CC (7) a method of treating an immune disorder ex vivo in a subject. (I) has  
CC virucide, antibacterial, antiparasitic, cytostatic and immunosuppressive  
CC activities, and can be used in gene therapy. The MHC class II compound  
CC (I) can be used for preparing a composition for treating immune  
CC disorders, e.g., viral infections, bacterial infections, parasitic  
CC infections, neoplastic disease, autoimmunity or toxicity. The present  
CC sequence represents a spaceholder molecule peptide, which can be used in  
XX an MHC class II compound from the present invention.  
XX Sequence 13 AA;  
XX

Query Match 88.0%; Score 44; DB 8; Length 13;  
 Best Local Similarity 91.7%; Pred. No. 1.9;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAA 12  
 |||||  
 Db 1 AAAAAAAAAA 12  
 |||||

RESULT 3  
 ADI42025  
 ID ADI42025 standard; protein; 233 AA.  
 XX  
 AC ADI42025;  
 XX  
 DT 22-APR-2004 (first entry)  
 XX  
 DE Plant transcription factor #231.  
 XX  
 KW transgenic; plant; enhanced tolerance to abiotic stress;  
 KW glyphosate tolerance; hormone sensitivity; disease resistance;  
 KW sugar sensing; flowering; flower structure; stem bifurcation;  
 KW branching pattern; apical dominance; trichome; stem morphology;  
 KW root growth; root hair; seed development; cell proliferation;  
 KW cell differentiation; premature senescence; necrosis; plant size;  
 KW leaf morphology; seed morphology; seed biochemistry; root anthocyanin;  
 KW plant anthocyanin; light response; shade avoidance; bioinformatic;  
 KW transcription factor; gene; ds.  
 XX  
 OS Oryza sativa.  
 XX  
 PN US2004019927-A1.  
 XX  
 XX 29-JAN-2004.  
 XX  
 XX 25-FEB-2003; 2003US-00374780.  
 XX  
 PR 18-APR-2001; 2001US-00837944.  
 XX  
 PA (SHER/) SHERMAN B K.  
 PA (RIEC/) RIECHMANN J L.  
 PA (JIAN/) JIANG C.  
 PA (HEAR/) HEARD J E.  
 PA (HAAK/) HAAKE V.  
 PA (CREE/) CREELMAN R A.  
 PA (RATC/) RATCLIFFE O.  
 PA (ADAM/) ADAM L J.  
 PA (REUB/) REUBER T L.  
 PA (KEDD/) KEDDIE J.  
 PA (BROU/) BROUN P E.  
 PA (PILG/) PILGRIM M L.  
 PA (DUBE/) DUBELL A N.  
 PA (PINE/) PINEDA O.  
 PA (YUGG/) YU G.  
 XX  
 PI Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V;  
 PI Creelman RA, Ratcliffe O, Adam LJ, Reuber TL, Keddie J, Broun PE;  
 PI Pilgrim ML, Dubell AN, Pineda O, Yu G;  
 XX  
 DR WPI; 2004-132245/13.  
 XX  
 XX New transgenic plant comprising a recombinant polynucleotide of any one  
 PT of more than 500 nucleotide sequences, useful in bioinformatic search  
 PT methods.  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 488; 435pp; English.  
 XX  
 CC The invention describes a transgenic plant comprising a recombinant  
 CC polynucleotide of any one of more than 500 nucleotide sequences fully  
 CC defined in the specification or its complement. The method of the  
 CC invention can be used to produced a plant having altered traits such as:  
 CC enhanced tolerance to abiotic stress; glyphosphate tolerance; hormone  
 CC sensitivity; disease resistance; sugar sensing; early or late flowering;

CC altered flower structure, change in stem bifurcations, altered branching  
 CC pattern, reduced apical dominance, reduced trichome density; lack of  
 CC trichomes; reduced ectopic trichome development; altered trichome  
 CC development; increase in trichome number; altered stem morphology;  
 CC increased root growth; increased root hairs; altered seed development;  
 CC altered cell proliferation or cell differentiation; rapid development;  
 CC premature senescence; increased necrosis; increase in seedling or plant  
 CC size; decreased plant size; leaf morphology; seed morphology; seed  
 CC biochemistry; increase in root anthocyanins; increase in plant  
 CC anthocyanins; or alteration in light response or shade avoidance. The  
 CC transgenic plant, polynucleotides and polypeptides are useful in  
 CC bioinformatic search methods. This is the amino acid sequence of a plant  
 CC transcription factor, and an orthologue of Arabidopsis thaliana  
 CC creation of a transgenic plant with altered traits.  
 XX  
 XX Sequence 233 AA;  
 SQ

Query Match 88.0%; Score 44; DB 8; Length 233;  
 Best Local Similarity 91.7%; Pred. No. 33;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAA 12  
 |||||  
 Db 18 AAAAAAAAAA 29  
 |||||

RESULT 4  
 ADO02575  
 ID ADO02575 standard; protein; 233 AA.  
 XX  
 AC ADO02575;  
 XX  
 DT 01-JUL-2004 (first entry)  
 XX  
 DE Thalecress transcription factor, Rice orthologue #1.  
 XX  
 KW Rice; transcription factor; plant; transgenic; abiotic stress;  
 KW cold tolerance; heat tolerance; drought; osmotic stress;  
 KW phosphate limitation; potassium limitation; nitrogen limitation;  
 KW hormone sensitivity; disease resistance; sugar sensing; seed germination;  
 KW flowering; inflorescence architectural change;  
 KW meristem cell differentiation; phyllotaxy; apical dominance;  
 KW trichome development; seed development; premature senescence;  
 KW delayed senescence; lethality; necrosis; plant size; leaf morphology;  
 KW seed morphology; secondary metabolism; light response; shade avoidance.  
 XX  
 OS Oryza sativa.  
 XX  
 PN US2004045049-A1.  
 XX  
 PD 04-MAR-2004.  
 XX  
 PF 10-APR-2003; 2003US-00412699.  
 XX  
 PR 13-SEP-1999; 99US-00394519.  
 PR 21-JAN-2000; 2000US-00489376.  
 PR 17-FEB-2000; 2000US-00506720.  
 PR 22-MAR-2000; 2000US-00532591.  
 PR 22-MAR-2000; 2000US-00533029.  
 PR 22-MAR-2000; 2000US-00533030.  
 PR 22-MAR-2000; 2000US-00533392.  
 PR 22-MAR-2000; 2000US-00533648.  
 PR 06-APR-2000; 2000WO-US009448.  
 PR 16-NOV-2000; 2000US-00713994.  
 PR 27-MAR-2001; 2001US-00819142.  
 PR 17-APR-2001; 2001US-00837444.  
 PR 30-JAN-2002; 2002US-00958131.  
 PR 14-JUN-2002; 2002US-00171468.  
 PR 09-AUG-2002; 2002US-00225066.  
 PR 09-AUG-2002; 2002US-00225067.  
 PR 09-AUG-2002; 2002US-00225068.  
 PR 17-DEC-2002; 2002US-0434166P.

25-FEB-2003; 2003US-00374780.

(ZHAN/) ZHANG J.  
(FROM/) FROMM M E.  
(HEAR/) HEARD J E.  
(RIEC/) RIECHMANN J L.  
(ADAM/) ADAM L J.  
(BROU/) BROUN P E.  
(PINE/) PINEDA O.  
(REUB/) REUBER T L.  
(KEDD/) KEDDIE J S.  
(YUGG/) YU G.  
(JIANG/) JIANG C.  
(SAMA/) SAMAHA R S.  
(PILG/) PILGRIM M L.  
(CREE/) CREELMAN R A.  
(DUBE/) DUBELL A N.  
(RATC/) RATCLIFFE O.  
(KUMI/) KUMIMOTO R.  
(SHER/) SHERMAN B K.

Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE, Pineda O, Reuber TL, Keddie JS, Yu G, Jiang C, Samaha RS, Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R, Sherman BK;

WPI; 2004-225755/21.

New transgenic plant, useful in developing phenotypes with altered or improved characteristics or traits.

Claim 1; SEQ ID NO 989; 213pp; English.

The invention relates to a transgenic plant comprising a recombinant polynucleotide having a polynucleotide sequence or its complementary sequence comprising a sequence encoding a polypeptide, that initiates transcription (i.e. a transcription factor) from Arabidopsis, Soybean, Rice, Rape or Corn, comprising any of the sequences appearing as ADO01588 - ADO03527 or ADO03530-AO03559. Also included are using a transgenic plant to grow a progeny plant, an expression cassette (comprising a constitutive, inducible or tissue-specific promoter and a recombinant polynucleotide described above), a host cell comprising the expression cassette, producing a modified plant having a modified trait, identifying a factor that is modulated by or interacts with a polypeptide encoded by the polynucleotide sequence and identifying at least one downstream polynucleotide sequence that is subject to a regulatory effect of any of the polypeptides encoded by the polynucleotide described above. The transgenic plant is useful for producing a plant that has an altered trait e.g. an enhanced tolerance to abiotic stress (increased tolerance to chilling, germination in cold conditions, freezing tolerance, tolerance to heat, tolerance to drought, tolerance to osmotic stress, tolerance to salt, tolerance to phosphate limitation, tolerance to potassium limitation, decreased sensitivity to nitrogen limitation), altered hormone sensitivity, reduced sensitivity to abscisic acid, an altered response to ethylene, disease resistance, altered susceptibility to Botrytis, altered susceptibility to Fusarium, altered susceptibility to Erysiphe, altered susceptibility to Pseudomonas syringae, altered susceptibility to Sclerotinia, altered sugar sensing, improved seed germination and seedling vigor, early flowering, late flowering, extended period of flowering, an inflorescence architectural change, a change in stem bifurcations, a lack of a shoot meristem, reduced meristem cell differentiation, altered phyllotaxy, altered branching pattern, reduced apical dominance, reduced trichome density, ectopic trichome development, altered trichome development, altered seed morphology, increased root growth, increased root hairs, altered seed development, altered cell proliferation/cell differentiation, premature senescence, delayed senescence, lethality, increased necrosis, an increase in seedling or plant size, decreased plant size, a change in leaf morphology, increased altered leaf development, increased leaf size and mass, glossy leaves, leaf cell expansion, change in seed morphology, altered seed coloration, increased seed size, decreased seed size, altered seed shape, change in leaf biochemistry, increased leaf wax, an alteration in leaf prenol lipid content, increased leaf insoluble sugars, decreased leaf insoluble

CC sugars, increased leaf anthocyanins, an alteration of leaf fatty acid  
CC content, an alteration of leaf glucosinolate content, change in seed  
CC biochemistry, an increase in seed oil content, decrease in seed oil  
CC content, increase in seed fatty acid content, decrease in seed fatty acid  
CC content, increase in seed protein content, decrease in seed protein  
CC content, alteration in seed prenol lipid content, increase in seed  
CC sterols, upregulation of genes involved in secondary metabolism, increase  
CC in root anthocyanins, increase in plant anthocyanins, and alteration in  
CC light response or shade avoidance. The present sequence represents an  
CC orthologue of a thalecress transcription factor isolated from Rice.  
XX  
XX  
SQ Sequence 233 AA;

Query Match 88.0%; Score 44; DB 8; Length 233;  
Best Local Similarity 91.7%; Pred. No. 33;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
|||  
Db 18 AAAAAAAAAA 29

RESULT 5  
ADO62214  
ID ADO62214 standard; protein; 233 AA.  
XX  
AC ADO62214;  
XX  
DT 15-JUL-2004 (first entry)  
XX  
DE Transcription factor G12 orthologous sequence, SEQ ID 681.  
XX  
KW Plant; transcription factor; transgenic plant; abiotic stress tolerance;  
KW osmotic stress tolerance; cold tolerance; heat tolerance;  
KW low nitrogen tolerance; low phosphate tolerance; fungal disease;  
KW glyphosate resistance; flowering; fertility; seed development.  
XX  
OS Oryza sativa.  
XX  
PN WO2004031349-A2.  
XX  
PD 15-APR-2004.  
XX  
PF 18-SEP-2003; 2003WO-US030292.  
XX  
PR 18-SEP-2002; 2002US-0411837P.  
PR 17-DEC-2002; 2002US-0434166P.  
PR 24-APR-2003; 2003US-0465809P.  
XX  
(MEND-) MENDEL BIOTECHNOLOGY INC.  
XX  
PI Jiang C, Heard JE, Ratcliffe O, Creelman RA, Adam LJ, Reuber TL;  
PI Riechmann JL, Haake V, Dubell AN, Keddie JS, Sherman BK;  
XX  
WPI; 2004-330163/30.

New recombinant polynucleotide encoding transcription factor polypeptides, useful for producing transgenic plants with advantageous properties compared to a reference plant.

Claim 1; SEQ ID NO 681; 510pp; English.

The present invention relates to novel plant transcription factor proteins (I) and nucleotide sequences (II) (ADO61534-AO63778). The sequences can be used to produce transgenic plants, which overexpress (II), where the transgenic plant has an altered trait as compared to a non-transgenic plant or wild-type plant. The transgenic plant comprises an altered trait selected from increased tolerance to abiotic stress, increased tolerance to osmotic stress, increased tolerance to heat, increased germination in cold, increased tolerance to freezing conditions, increased tolerance to low nitrogen conditions, increased tolerance to low phosphate conditions, increased tolerance to disease, including

CC fungal disease and particularly Erysiphe, Fusarium and Botrytis,  
CC increased tolerance to multiple fungal pathogens, increased resistance to  
CC glyphosate, increased sensitivity to ABA, reduced sensitivity to ABA,  
CC increased sensitivity to ACC, altered sugar sensing, increased tolerance  
CC to sugars, altered carbon/nitrogen sensing, early flowering, late  
CC flowering, altered flower structure, loss of flower determinacy, reduced  
CC fertility, altered shoot meristem development, altered branching pattern  
CC altered stem morphology, altered vascular tissue structure, reduced  
CC apical dominance, altered trichome density, altered trichome development,  
CC altered trichome structure, altered root development, altered shade  
CC avoidance, altered seed development, altered seed ripening, altered seed  
CC germination, slow growth, fast growth, altered cell differentiation,  
CC altered cell proliferation, altered cell expansion, altered phase change,  
CC altered senescence, abnormal embryo development, altered programmed cell  
CC death, lethality when overexpressed, altered necrosis patterns, increased  
CC plant size, increased biomass, large seedlings, dwarfed plants, dark  
CC green leaves, change in leaf shape, increased leaf size and mass, light  
CC green or gray leaves, glossy leaves, altered abaxial/adaxial polarity,  
CC altered seed coloration, altered seed size, altered seed shape, large  
CC seed, increased leaf wax, increased leaf fatty acids, altered seed oil  
CC content, altered seed protein content, altered seedprenyl content,  
CC altered leaf prenyl lipid content, increased anthocyanin levels, and  
CC decreased anthocyanin levels. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published pct sequences.

Sequence 233 AA:

Query Match	88.0%	Score 44;	DB 8;	Length 233;
Best Local Similarity	91.7%	Pred. No. 33;		
Matches 11;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			

QY 1 AAAAAAAAAA 12  
||| ||| ||| |||  
Db 18 AAAAAAAAAA 29

## RESULT 6

AEF30018  
ID AEF30018 standard: protein: 233 AA.

AC AEF30018:

XX  
DT 23-MAR-2006 (first entry)

DE CeresClone681222 protein homolog SEO ID NO:2218. XX

AA plant; transgenic plant; crop improvement; abiotic stress tolerance;  
KW plant growth regulation.

XX  
OS  
parthenium argentatum.

AA  
PN  
WO2006004955-A2.

12-JAN-2006.

30-JUN-2005: 2005WO-US023326.

PR 30-JUN-2004; 2004US-0583621P.

PR 30-JUN-2004; 2004US-0584800P;  
PR 30-JUN-2004; 2004US-0584829P;

XX  
PA (CERE-) CERES INC.

XX  
PT  
Alexandrov N. Bro

XX  
DB WPT. 2006-090599/09

XX  
PT  
New isolated nucleic

PT characteristics and the polypeptide it encodes, useful for making transgenic plants with improved characteristics.

XX

Claim 1: SEO ID NO 2218; 612pp; English.

The invention relates to an isolated nucleic acid molecule modifying plant phenotypes and characteristics, comprising a nucleotide sequence that encodes an amino acid sequence exhibiting at least 85% sequence identity to an amino acid sequence in the sequence listing or in the ortholog alignments of Figure 1, a nucleic acid, which is a complement of (a), a nucleic acid, which is the reverse of the nucleotide sequence in (a) (such that the reverse nucleotide sequence has a sequence order which is the reverse of the sequence order of (a)) or a nucleic acid capable of hybridizing (a-c), under conditions that permit formation of a nucleic acid duplex at a temperature of 40-48 degrees C below the melting temperature of the nucleic acid duplex. Also included are a vector construct (comprising a first nucleic acid having a regulatory sequence capable of causing transcription and/or translation in a plant, operably linked to a second nucleic acid having the sequence of the isolated nucleic acid molecule), a host cell comprising the isolated nucleic acid molecule that is flanked by exogenous sequence, a host cell comprising the vector construct, an isolated polypeptide comprising an amino acid sequence exhibiting at least 85% sequence identity to those cited above, introducing an isolated nucleic acid into a host cell, transforming a host cell, detecting a nucleic acid in a sample, a host cell or organism comprising the nucleic acid molecule, a plant generated from the plant cell or seed, a plant/plant cell, plant material or seed) comprising a nucleic acid molecule (where the plant has improved characteristics as compared to a wild type plant), improving plant characteristics in a plant comprising transforming the plant with the nucleic acid sequence, and a transgenic plant having a gene construct (comprising the nucleic acid encoding a component operably linked to a plant promoter so that the component is ectopically overexpressed in the transgenic plant). The transgenic plant exhibits faster rate of growth, greater fresh or dry weight of maturation, greater fruit or seed yield, higher tolerance to pH, higher tolerance to low phosphate concentration, or higher tolerance to low nitrogen concentration than a progenitor plant, which does not contain the progenitor construct, when the transgenic plant and progenitor plant are cultivated under identical environmental conditions, where the component is any one of the polypeptides cited above. The nucleic acid molecules are useful for producing transgenic plants with improved characteristics. The present sequence is an ortholog of a protein encoded by a plant nucleic acid (cDNA) of the invention.

Sequence 233 AA;

Query Match 88.0%; Score 44; DB 10; Length 233;

Query Match 88.0%; score 44; DE  
Best Local Similarity 91.7%; Pred. No. 33;

BEST LOCAL SIMILARITY 91.7%; PRED. NO. 33;  
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AAAAAAAAAA 12

## RESULT 7

RESOLUT  
ADR98917

ID ADR98917 standard; protein; 278 AA.

AC ADR98917;

DT 18-NOV-2004 (first entry)

DE Lung specific gene splice variant encoded protein #89.

XX  
KW cytostatic; gene therapy; vaccine; lung; diagnosis; cancer;

KW non-cancerous lung disease; lung  
transgenic animal: splice variant.

XX OS Homo sapiens.

XX  
DN  
W02004074430-A2XX  
02-SEP-2004XX  
PF  
08-DEC-2003: 2003WO-IIS038896-

XX 06-DEC-2002; 2002US-0431307P.  
 PR 06-DEC-2002; 2002US-0431510P.  
 PR 06-DEC-2002; 2002US-0431516P.  
 XX (DIAD-) DIADEXUS INC.  
 XX Macina RA, Turner LR, Sun Y, Liu S;  
 PI WPI; 2004-635553/61.  
 DR N-PSDB; ADR98780.  
 XX New isolated human lung specific nucleic acid molecule, useful for  
 PT identifying, diagnosing, monitoring, staging, imaging and treating lung  
 PT cancer and non-cancerous diseases of the lung.  
 XX Claim 1; SEQ ID NO 202; 542pp; English.  
 XX The invention relates to a new isolated lung specific nucleic acid  
 CC molecule (I) comprising any of 113 fully defined nucleotide sequences  
 CC given in the specification, their encoded protein sequences, sequences  
 CC selectively hybridizing to the nucleotide sequences or a sequence having  
 CC at least 60% identity to the nucleotide sequences. The methods and  
 CC compositions of the present invention are useful for identifying,  
 CC diagnosing, monitoring, staging, imaging and treating lung cancer and non  
 CC -cancerous diseases of the lung. They are also used for identifying lung  
 CC tissue, monitoring and identifying and/or designing antagonists of the  
 CC polypeptide of the invention, gene therapy, production of transgenic  
 CC animals and production of engineered lung tissue for treatment and  
 CC research. Lung specific genes (LSGs) were identified by a systematic  
 CC analysis of gene expression data in the LIFESEQ Gold database using the  
 CC data mining software package candidate lead automatic search program  
 CC (CLASP). Genes were grouped into gene bins where each bin is a cluster of  
 CC sequences grouped together where they share a common contig.  
 CC Differentially expressed tissue-specific genes were selected based on the  
 CC percentage level in the targeted tissue versus all the other tissues. The  
 CC expression levels for each gene in libraries of normal tissues or non-  
 CC tumour tissues from cancer patients were compared with the expression  
 CC levels in tissue libraries associated with tumour or disease. This  
 CC sequence represents a protein of the invention.  
 XX Sequence 278 AA;  
 SQ Query Match 88.0%; Score 44; DB 8; Length 278;  
 Best Local Similarity 91.7%; Pred. No. 40;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAA 12  
 |||||  
 Db 1 AAAAAAAAAA 12  
 RESULT 8  
 ADY09213  
 ID ADY09213 standard; protein; 646 AA.  
 XX AC ADY09213;  
 XX 21-APR-2005 (first entry)  
 DT Plant full length insert polypeptide seqid 65028.  
 DE plant protectant; plant growth regulator; gene therapy; plant;  
 KW recombinant DNA construct; physical array; plant breeding marker;  
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
 KW growth rate; cell cycle pathway; disease resistance;  
 KW galactomannan production; lignin production; plant growth regulator;  
 KW yield; plant growth; plant development; seed oil; protein yield;  
 KW protein content.  
 XX Unidentified.  
 OS  
 XX

PN US2004034888-A1.  
 XX 19-FEB-2004.  
 PD 28-APR-2003; 2003US-00425114.  
 XX 06-MAY-1999; 99US-00304517.  
 PR 05-NOV-2001; 2001US-00985678.  
 XX (LIUJ/) LIU J.  
 PA (ZHOU/) ZHOU Y.  
 PA (KOVA/) KOVALIC D K.  
 PA (SCRE/) SCREEN S E.  
 PA (TABAS/) TABASKA J E.  
 PA (CAOY/) CAO Y.  
 XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
 PI WPI; 2004-180133/17.  
 DR New recombinant DNA construct, useful for improving plant tolerance to  
 XX cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
 PT pests, for conferring increased resistance to plant disease, or for  
 PT improving yield.  
 XX Claim 1; SEQ ID NO 65028; 15pp; English.  
 PS The invention describes a recombinant DNA construct comprising a  
 CC polynucleotide consisting of a sequence encoding an amino acid sequence  
 CC available in electronic form from the US patent office at  
 CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
 CC of the invention are also useful in physical arrays of molecules and as  
 CC plant breeding markers. The recombinant DNA construct is useful for  
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
 CC plant cells by modification of the cell cycle pathway, for conferring  
 CC increased resistance to plant disease, for producing galactomannan,  
 CC lignin or plant growth regulators, for increasing the rate of homologous  
 CC recombination in plants, for improving yield by modification of  
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC or by providing improved plant growth and development under at least one  
 CC stress condition or for modifying seed oil or protein yield and/or  
 CC content. This is the amino acid sequence of a plant full length insert  
 CC polypeptide that can be used in the recombinant DNA construct of the  
 CC invention.  
 XX Sequence 646 AA;  
 SQ Query Match 88.0%; Score 44; DB 8; Length 646;  
 Best Local Similarity 91.7%; Pred. No. 92;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AAAAAAAAAA 12  
 |||||  
 Db 4 AAAAAAAAAA 15  
 RESULT 9  
 ABB67142  
 ID ABB67142 standard; protein; 526 AA.  
 XX AC ABB67142;  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster polypeptide SEQ ID NO 28218.  
 DE Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmacological.  
 KW Drosophila melanogaster.  
 XX OS  
 XX WO200171042-A2.  
 PN



XX PD 27-SEP-2001.  
 XX XX 23-MAR-2001; 2001WO-US009231.  
 XX PF  
 XX XX 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX XX (PEKE ) PE CORP NY.  
 XX PA Venter JC, Adams M, Li PWD, Myers EW;  
 XX PI WPI; 2001-656860/75.  
 XX XX N-PSDB; ABL11245.  
 DR  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 28218; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signaling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 526 AA;  
 PS  
 XX Query Match 86.0%; Score 43; DB 4; Length 526;  
 CC Best Local Similarity 83.3%; Pred. No. 1e+02;  
 CC Matches 10; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 AAAAAAAAAA 12  
 DB 105 AAAAAAAAAA 116  
 ||:|||||:|  
 RESULT 10  
 ABB64188  
 ID ABB64188 standard; protein; 2857 AA.  
 XX  
 AC ABB64188;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX Drosophila melanogaster polypeptide SEQ ID NO 19356.  
 DE  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW  
 XX Drosophila melanogaster.  
 OS  
 XX WO200171042-A2.  
 PN  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US009231.  
 FF  
 XX 23-MAR-2000; 2000US-0191637P.  
 XX PR 11-JUL-2000; 2000US-00614150.  
 XX PR  
 XX (PEKE ) PE CORP NY.  
 PA Venter JC, Adams M, Li PWD, Myers EW;  
 XX PI WPI; 2001-656860/75.  
 XX XX N-PSDB; ABL08291.  
 DR

XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 19356; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2857 AA;  
 PS  
 XX Query Match 84.0%; Score 42; DB 4; Length 2857;  
 CC Best Local Similarity 90.9%; Pred. No. 7.9e+02;  
 CC Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 AAAAAAAAAA 11  
 DB 135 AAAAAAAAAA 145  
 |||||||  
 RESULT 11  
 ABB06570  
 ID ABB06570 standard; protein; 129 AA.  
 XX  
 AC ABB06570;  
 AC  
 XX 24-JUN-2002 (first entry)  
 DT  
 XX Human ORFX protein sequence SEQ ID NO:13122.  
 DE  
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KW myasthenia gravis.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200192523-A2.  
 FN  
 XX 06-DEC-2001.  
 PD  
 XX 29-MAY-2001; 2001WO-US010836.  
 PF  
 XX 30-MAY-2000; 2000US-0206132P.  
 PR  
 XX 29-AUG-2000; 2000US-0228716P.  
 PR  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX Shimkets RA, Leach MD;  
 PI  
 XX WPI; 2002-106308/14.  
 DR  
 XX N-PSDB; ABB22322.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.  
 XX  
 PS Disclosure; SEQ ID NO 13122; 1037pp; English.  
 XX  
 XX The present invention describes substantially purified human proteins

CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
CC in the specification). ABN1562 to ABN2752 encode the human ORFX  
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for  
CC treating or preventing a pathology associated with an ORFX-associated  
CC disorder in humans, and in the manufacture of a medicament for treating a  
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
CC sequences can be used in gene therapy. ORFX sequences can be used in the  
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
CC storage disease, various immune deficiencies and disorders, infectious  
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
CC bone degenerative disorders, or periodontal disease, and for gut  
CC protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues and conditions resulting from  
CC systemic cytokine damage. N.B. The sequence data for this patent did not  
CC form part of the printed specification, but was obtained in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 129 AA;

Query Match 82.0%; Score 41; DB 5; Length 129;  
Best Local Similarity 83.3%; Pred. No. 51;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAMAAAAAAWA 12  
|:|||||  
Db 11 AALAAAAAAWA 22

## RESULT 12

AAG32713  
ID AAG32713 standard; protein; 156 AA.

XX AC AAG32713;

XX DT 17-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 39519.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; Genetic mapping; gene expression control; promoter;  
XX termination sequence; corn.

XX OS Zea mays subsp. mays.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137223P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.



CC hybridizing (a-c), under conditions that permit formation of a nucleic  
CC acid duplex at a temperature of 40-48 degrees C below the melting  
CC temperature of the nucleic acid duplex. Also included are a vector  
CC construct (comprising a first nucleic acid having a regulatory sequence  
CC capable of causing transcription and/or translation in a plant, operably  
CC linked to a second nucleic acid having the sequence of the isolated  
CC nucleic acid molecule), a host cell comprising the isolated nucleic acid  
CC molecule that is flanked by exogenous sequence, a host cell comprising  
CC the vector construct, an isolated polypeptide comprising an amino acid  
CC sequence exhibiting at least 85% sequence identity to those cited above,  
CC introducing an isolated nucleic acid into a host cell, transforming a  
CC host cell, detecting a nucleic acid in a sample, a host cell or organism  
CC comprising the nucleic acid molecule, a plant generated from the plant  
CC cell or seed, a plant/plant cell, plant material (or seed) comprising the  
CC nucleic acid molecule (where the plant has improved characteristics as  
CC compared to a wild type plant), improving plant characteristics in a  
CC plant comprising transforming the plant with the nucleic acid sequence,  
CC and a transgenic plant having a gene construct (comprising the nucleic  
CC encoding a component operably linked to a plant promoter so that the  
CC component is ectopically overexpressed in the transgenic plant). The  
CC transgenic plant exhibits faster rate of growth, greater fresh of dry  
CC weight of maturation, greater fruit or seed yield, higher tolerance to  
CC pH, higher tolerance to low phosphate concentration, or higher tolerance  
CC to low nitrogen concentration than a progenitor plant, which does not  
CC contain the progenitor construct, when the transgenic plant and  
CC progenitor plant are cultivated under identical environmental conditions,  
CC where the component is any one of the polypeptides cited above. The  
CC nucleic acid molecules are useful for producing transgenic plants with  
CC improved characteristics. The present sequence is an ortholog of a  
CC protein encoded by a plant nucleic acid (cDNA) of the invention.

XX SQ Sequence 243 AA;

Query Match 82.0%; Score 41; DB 10; Length 243;  
Best Local Similarity 83.3%; Pred. No. 96;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAAAAAAAAA 12  
Db 18 SAAAAAAAAA 29  
: |||||

RESULT 14  
ADI60138  
ID ADI60138 standard; protein; 276 AA.

XX AC ADI60138;

XX DT 15-APR-2004 (first entry)

XX DE Secreted polypeptide #22.

XX osteopathic; vulnery; cytostatic; gene therapy; diagnosis; forensics;  
XX gene mapping; mutation identification; biodiversity; chromosome marker;  
XX immune response; myeloid cell disorder; lymphoid cell disorder;  
XX bone cartilage; tendon; ligament; nerve tissue growth; wound healing;  
XX burns; incision; ulcer; cancer.

XX OS Homo sapiens.

XX PN WO2003025142-A2.

XX PD 27-MAR-2003.

XX PF 18-SEP-2002; 2002WO-US029636.

XX PR 18-SEP-2001; 2001US-0323349P.

XX PR 16-SEP-2002; 2002US-00323349.

XX PA (HYSE-) HYSEQ INC.

XX Tang YF, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT;

XX WPI; 2003-354601/33.  
DR N-PSDB; ADI60483.

XX New polynucleotides and secreted proteins, useful for treating myeloid or  
PT lymphoid cell disorders, in bone cartilage, tendon, ligament and nerve  
PT tissue growth or regeneration, in wound healing, and in tissue repair and  
PT replacement.

XX Claim 20; SEQ ID NO 173; 243pp; English.

XX The invention relates to novel isolated polynucleotides or a sequence  
CC encoding a polypeptide with biological activity, where the polynucleotide  
CC hybridizes to the polynucleotide under stringent hybridization conditions  
CC or has greater than 99% sequence identity with the polynucleotide. The  
CC polynucleotides and polypeptides are useful in the diagnostics, forensics,  
CC gene mapping, identification of mutations responsible for genetic  
CC disorders and other traits, to assess biodiversity, as nutritional  
CC sources or supplements. The polynucleotides may also be used as molecular  
CC weight markers, chromosome markers or map related gene positions, or as  
CC an antigen to raise anti-DNA antibodies or elicit immune response. The  
CC polypeptides are useful for raising antibodies, as markers for tissues in  
CC which the corresponding polypeptide is expressed, for re-engineering  
CC damaged or diseased tissues, for treating myeloid or lymphoid cell  
CC disorders, in bone cartilage, tendon, ligament and/or nerve tissue growth  
CC or regeneration, in wound healing, in tissue repair and replacement, in  
CC healing of burns, incisions and ulcers, and in treating cancer. This  
CC sequence corresponds to a protein sequence of the invention.

XX SQ Sequence 276 AA;

Query Match 82.0%; Score 41; DB 7; Length 276;  
Best Local Similarity 75.0%; Pred. NO. 1.1e+02;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAAAAAAAAA 12  
Db 5 AALGAAAAAALA 16  
||:|||||

RESULT 15  
ABB69217

ID ABB69217 standard; protein; 328 AA.

XX AC ABB69217;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 34443.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX DR N-PSDB; ABL13320.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

PS Disclosure; SEQ ID NO 34443; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 328 AA;

Query Match 82.0%; Score 41; DB 4; Length 328;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 10

Db 98 AAAAAAAAAA 107

|||||

RESULT 16

ID ADX78446 standard; protein; 333 AA.

XX

AC ADX78446;

XX

XX 23-MAR-2006 (revised)

DT 21-APR-2005 (first entry)

XX

XX Plant full length insert polypeptide seqid 47812.

DE

XX plant protectant; plant growth regulant; gene therapy; plant;

KW recombinant DNA construct; physical array; plant breeding marker;

KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;

KW extreme osmotic condition; pathogen tolerance; pest tolerance;

KW growth rate; cell cycle pathway; disease resistance;

KW galactomannan production; lignin production; plant growth regulator;

KW yield; plant growth; plant development; seed oil; protein yield;

KW protein content.

XX

OS Zea mays.

XX

XX US2004034888-A1.

XX

XX 19-FEB-2004.

XX

XX 28-APR-2003; 2003US-00425114.

XX

XX 06-MAY-1999; 99US-00304517.

XX

XX 05-NOV-2001; 2001US-00985678.

XX

XX (LIUJ/) LIU J.

XX (ZHOU/) ZHOU Y.

XX (KOVA/) KOVALIC D K.

XX (SCRE/) SCREEN S E.

XX (TABA/) TABASKA J E.

XX (CAOY/) CAO Y.

XX

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

PI WPI; 2004-180133/17.

XX

XX New recombinant DNA construct, useful for improving plant tolerance to

PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or

PT pests, for conferring increased resistance to plant disease, or for

PT improving yield.

XX

PS Claim 1; SEQ ID NO 47812; 15pp; English.

XX

XX The invention describes a recombinant DNA construct comprising a

CC polynucleotide consisting of a sequence encoding an amino acid sequence

CC available in electronic form from the US patent office at

CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide

CC of the invention are also useful in physical arrays of molecules and as

CC plant breeding markers. The recombinant DNA construct is useful for

CC improving plant tolerance to cold, heat, drought, herbicides, extreme

CC osmotic conditions, pathogens or pests, for manipulating growth rate in

CC plant cells by modification of the cell cycle pathway, for conferring

CC increased resistance to plant disease, for producing galactomannan,

CC lignin or plant growth regulators, for increasing the rate of homologous

CC recombination in plants, for improving yield by modification of

CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake

CC or by providing improved plant growth and development under at least one

CC stress condition or for modifying seed oil or protein yield and/or

CC content. This is the amino acid sequence of a plant full length insert

CC polypeptide that can be used in the recombinant DNA construct of the

CC invention.

CC

CC Revised record issued on 23-MAR-2006 : Corrected organism line

XX

XX Sequence 333 AA;

SQ

Query Match 82.0%; Score 41; DB 8; Length 333;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12

Db 157 SAMAAAAA 168

|||||

RESULT 17

ID ADX96203 standard; protein; 429 AA.

XX

XX ADX96203;

XX

XX 21-APR-2005 (first entry)

DT

XX

XX Plant full length insert polypeptide seqid 58867.

XX

XX plant protectant; plant growth regulant; gene therapy; plant;

KW recombinant DNA construct; physical array; plant breeding marker;

KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;

KW extreme osmotic condition; pathogen tolerance; pest tolerance;

KW growth rate; cell cycle pathway; disease resistance;

KW galactomannan production; lignin production; plant growth regulator;

KW yield; plant growth; plant development; seed oil; protein yield;

KW protein content.

XX

XX Unidentified.

XX

XX US2004034888-A1.

XX

XX 19-FEB-2004.

XX

XX 28-APR-2003; 2003US-00425114.

XX

XX 06-MAY-1999; 99US-00304517.

XX

XX 05-NOV-2001; 2001US-00985678.

XX

XX (LIUJ/) LIU J.

XX (ZHOU/) ZHOU Y.

XX (KOVA/) KOVALIC D K.

XX (SCRE/) SCREEN S E.

XX (TABA/) TABASKA J E.

XX (CAOY/) CAO Y.

XX

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

PI

```
DR WPI; 2004-180133/17.
XX
PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 5867; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?docID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 429 AA;
Query Match 82.0%; Score 41; DB 8; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAA 10
Db 20 AAAAAAAAAA 29
RESULT 18
AAE05813
ID AAE05813 standard; protein; 443 AA.
XX
AC AAE05813;
XX
DT 24-SEP-2001 (first entry)
XX
DE Human small cell lung cancer associated protein, SOX3.
XX
KW Human; small cell lung cancer; therapy; hCAAP; nucleic acid; NA;
KW melanoma; cancer; colon; breast; head; neck; transitional cancer;
KW leiomyosarcoma; synovial sarcoma; cytostatic; SOX3.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 137..215
FT /note= "DNA-binding HMG domain"
XX
PN WO200153349-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US002015.
XX
PR 21-JAN-2000; 2000US-00489101.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PA (SLOK ) SLOAN KETTERING INST CANCER RES.
XX (CORR ) CORNELL RES FOUND INC.
XX
PI Stockert E, Scanlan MJ, Jager D, Old LJ, Gure AO, Chen Y;
XX
```

---

```
DR WPI; 2001-457597/49.
DR N-PSDB; AAD11118.
XX
PT Isolated polypeptide, used to treat or prognose a disorder characterized
PT by expression of a hCAAP e.g. cancer, is encoded by an isolated nucleic
PT acid comprising an NA Group 3 or 4 molecule.
XX
PS Disclosure; Fig 1; 152pp; English.
XX
CC The invention relates to nucleic acids and encoded polypeptides which are
CC cancer associated antigens expressed in patients afflicted with small
CC cell lung cancer. The molecules provided by the invention can be used in
CC the diagnosis, monitoring, research or treatment of conditions
CC characterised by the expression of one or more cancer associated
CC antigens. The polypeptide is used to treat a disorder characterised by
CC expression of a hCAAP, and determine regression, progression or onset of
CC a condition characterised by expression of an abnormal amount of a
CC protein encoded by a nucleic acid (NA) Group 1 molecule. The disorders
CC are small and non-small cell lung cancer, melanoma, colon, breast, head
CC and neck, or transitional cancer, leiomyosarcoma or synovial sarcoma. The
CC present sequence is human SOX3 protein, encoded by a small cell lung
CC cancer associated gene designated as NY-SCLC-9
XX
SQ Sequence 443 AA;
Query Match 82.0%; Score 41; DB 4; Length 443;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAAAAAAAAA 12
Db 352 AAAAAAAAAA 363
RESULT 19
ABB71564
ID ABB71564 standard; protein; 524 AA.
XX
AC ABB71564;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 41484.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL15667.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 41484; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
```

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
CC AB872072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 524 AA;  
Query Match 82.0%; Score 41; DB 4; Length 524;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02; Indels 0; Gaps 0;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
Db 163 AAQAAAAAAAAA 174  
RESULT 20  
AAO20499  
ID AAO20499 standard; protein; 593 AA.  
XX  
AC AAO20499;  
XX  
DT 27-JUN-2002 (first entry)  
XX  
DE Protein of APP related human homologue hCP41313 #1.  
XX  
KW Neuroprotective; nootropic; transgenic fly; Alzheimer's disease; Abeta;  
KW amyloid precursor protein; tissue-specific expression control; human APP;  
KW APP pathway modulator; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FN WO200226820-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 01-OCT-2001; 2001WO-EP011345.  
XX  
PR 29-SEP-2000; 2000US-0236893P.  
PR 14-JUN-2001; 2001US-0298309P.  
XX  
PA (NOVS ) NOVARTIS AG.  
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
XX  
PI Cohen D, Dengler UJ, Finelli AL, Freuler F, Konsolaki M;  
PI Reinhardt MWHM, Zusman S;  
XX  
DR WPI; 2002-315796/35.  
DR N-PSDB; AAK99393.  
XX  
PT New transgenic fly, containing DNA encoding an Abeta portion of human  
PT APP, useful for identifying agents which modulate the APP pathway and  
PT which can be used to treat Alzheimer's disease.  
XX  
PS Claim 25; Page 90-91; 129pp; English.  
XX  
CC The invention relates to a transgenic fly whose genome comprises DNA  
CC encoding a polypeptide having the Abeta portion of human amyloid  
CC precursor protein (APP), fused to a signal sequence. The DNA sequence  
CC encodes a 123 (Abeta40) or 129 (Abeta42) amino acid sequence, given in  
CC the specification. The DNA sequence is operably linked to a tissue-  
CC specific expression control sequence. Expression of the sequence gives  
CC the fly an altered phenotype. The purpose of the invention is for  
CC identifying agents that inhibit or promote the expression and/or function  
CC of genes or encoded polypeptides which modify the APP pathway. The agent  
CC is a compound, triple helix DNA, antisense oligonucleotide, double  
CC stranded RNA molecule, ribozyme, or particularly an antibody. It is used  
CC to treat conditions such as Alzheimer's disease. The agent can be used as  
CC an APP pathway modulator or in gene therapy. This sequence represents the

CC protein of the APP related human homologue hCP41313 #1  
XX Sequence 593 AA;  
SQ  
Query Match 82.0%; Score 41; DB 5; Length 593;  
Best Local Similarity 75.0%; Pred. No. 2.3e+02; Indels 0; Gaps 0;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
Db 508 SSMASASAAAAA 519  
RESULT 21  
AAO20500  
ID AAO20500 standard; protein; 645 AA.  
XX  
AC AAO20500;  
XX  
DT 27-JUN-2002 (first entry)  
XX  
DE Protein of APP related human homologue hCP41313 #2.  
XX  
KW Neuroprotective; nootropic; transgenic fly; Alzheimer's disease; Abeta;  
KW amyloid precursor protein; tissue-specific expression control; human APP;  
KW APP pathway modulator; gene therapy.  
XX  
OS Homo sapiens.  
XX  
FN WO200226820-A2.  
XX  
PD 04-APR-2002.  
XX  
PF 01-OCT-2001; 2001WO-EP011345.  
XX  
PR 29-SEP-2000; 2000US-0236893P.  
PR 14-JUN-2001; 2001US-0298309P.  
XX  
PA (NOVS ) NOVARTIS AG.  
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
XX  
PI Cohen D, Dengler UJ, Finelli AL, Freuler F, Konsolaki M;  
PI Reinhardt MWHM, Zusman S;  
XX  
DR WPI; 2002-315796/35.  
DR N-PSDB; AAK99394.  
XX  
PT New transgenic fly, containing DNA encoding an Abeta portion of human  
PT APP, useful for identifying agents which modulate the APP pathway and  
PT which can be used to treat Alzheimer's disease.  
XX  
PS Claim 25; Page 92-93; 129pp; English.  
XX  
CC The invention relates to a transgenic fly whose genome comprises DNA  
CC encoding a polypeptide having the Abeta portion of human amyloid  
CC precursor protein (APP), fused to a signal sequence. The DNA sequence  
CC encodes a 123 (Abeta40) or 129 (Abeta42) amino acid sequence, given in  
CC the specification. The DNA sequence is operably linked to a tissue-  
CC specific expression control sequence. Expression of the sequence gives  
CC the fly an altered phenotype. The purpose of the invention is for  
CC identifying agents that inhibit or promote the expression and/or function  
CC of genes or encoded polypeptides which modify the APP pathway. The agent  
CC is a compound, triple helix DNA, antisense oligonucleotide, double  
CC stranded RNA molecule, ribozyme, or particularly an antibody. It is used  
CC to treat conditions such as Alzheimer's disease. The agent can be used as  
CC an APP pathway modulator or in gene therapy. This sequence represents the  
CC protein of the APP related human homologue hCP41313 #2  
XX  
SQ Sequence 645 AA;  
Query Match 82.0%; Score 41; DB 5; Length 645;  
Best Local Similarity 75.0%; Pred. No. 2.5e+02; Indels 0; Gaps 0;  
Matches 9; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
QY      1 AAAAAAAAAAAAA 12
DB      558 SSMASAAAAA 569

RESULT 22
ABM87699
ID ABM87699 standard; protein; 648 AA.
XX
AC ABM87699;
XX
XX 02-JUN-2005 (first entry)
XX
DE Rice abiotic stress responsive polypeptide SEQ ID NO:5945.
XX
KW abiotic stress tolerance; transgenic plant; plant; cereal; agriculture.
XX
OS Oryza sativa.
XX
PN WO2003008540-A2.
XX
XX 30-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019668.
XX
PF 22-JUN-2001; 2001US-0300112P.
PR 24-AUG-2001; 2001US-0314662P.
PR 26-SEP-2001; 2001US-0325277P.
PR 21-NOV-2001; 2001US-0332132P.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
PI Moughamer T, Provart N, Riecke D, Zhu T;
XX WPI; 2003-248011/24.
XX
XX New stress-responsive nucleic acid, useful for altering the
PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
PT stress, salt stress or osmotic stress.
XX
XX Claim 1; SEQ ID NO 5945; 89pp; English.
XX
XX The invention relates to novel abiotic stress responsive polynucleotides
CC and polypeptides. Also disclosed are vectors, expression cassettes, host
CC cells, and plants containing such polynucleotides. Also disclosed are
CC methods for using the polynucleotides and polypeptides to alter the
CC responsiveness of a plant to abiotic stress. The invention is useful in
CC agriculture. The nucleic acid is useful for determining whether a test
CC plant has been exposed to an abiotic stress condition. It is also useful
CC for selecting an agent that alters abiotic stress regulated
CC polynucleotide expression in a plant cell, and to identify a homolog or
CC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
CC molecule and the polypeptide encoded by it are useful in altering the
CC responsiveness of a plant to an abiotic stress, such as cold stress, salt
CC stress, osmotic stress or any of their combinations. The present sequence
CC is used in the exemplification of the invention
XX
SQ Sequence 648 AA;
Query Match 82.0%; Score 41; DB 7; Length 648;
Best Local Similarity 83.3%; Pred. No. 2.6e+02; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      1 AAAAAAAAAAAAA 12
DB      553 AAAAAAAAAAALA 564

RESULT 23
ABG23551
ID ABG23551 standard; protein; 791 AA.
XX
XX ABG23551 standard; protein; 791 AA.
XX
XX Claim 20; SEQ ID NO 53910; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 791 AA;
Query Match 82.0%; Score 41; DB 4; Length 791;
Best Local Similarity 83.3%; Pred. No. 3.1e+02; Mismatches 1; Indels 0; Gaps 0;
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      1 AAAAAAAAAAAAA 12
DB      496 AAAAAAAAAAAMS 507

RESULT 24
ADD02808
ID ADD02808 standard; peptide; 16 AA.
XX
XX ADD02808;
XX
```



XX 01-JAN-2004 (first entry)  
XX SCD28 molecule related motif peptide SEQ ID NO:7.  
XX superantigen binding site; CD28; superantigen; antibacterial;  
XX gene therapy; superantigen-related disorder; epitope.  
XX Synthetic.  
OS Homo sapiens.  
XX WO2003084995-A2.  
XX 16-OCT-2003.  
XX 03-APR-2003; 2003WO-IL000278.  
XX 04-APR-2002; 2002IL-00148993.  
XX (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.  
XX Kaempfer R, Arad G;  
XX WPI; 2003-804296/75.  
XX New superantigen binding site within the CD28 molecule, useful for  
PT preparing a pharmaceutical composition for treating superantigen-related  
PT disorders caused by Staphylococcus aureus or Streptococcus pyogenes.  
XX Example 8; SEQ ID NO 7; 160pp; English.  
XX The present invention describes a superantigen binding site within the  
XX CD28 molecule that specifically and directly binds to a superantigen.  
XX Also described: (1) a method for treating a superantigen-related disorder  
CC in a mammalian subject; (2) a method of inhibiting pyrogenic exotoxin-  
CC mediated activation of Th1-lymphocytes and protecting against toxic shock  
CC induced by a pyrogenic exotoxin or by a mixture of pyrogenic exotoxins,  
CC in a subject; (3) a method of eliciting protective immunity against toxic  
CC shock induced by a pyrogenic exotoxin in a subject; (4) a substance that  
CC inhibits the binding of a superantigen to a superantigen binding site in  
CC CD28; (5) a pharmaceutical composition for treating or preventing  
CC superantigen-related disorders comprising the substance that inhibits the  
CC direct interaction between CD28 molecule and the pyrogenic exotoxin,  
CC which leads to antagonising of toxin-mediated activation of Th1  
CC lymphocytes; (6) an isolated and purified peptide having an amino acid  
CC sequence homologous to an amino acid sequence comprised within a  
CC superantigen binding site within the CD28 molecule; (7) a method of  
CC screening for a test substance which specifically binds to the CD28  
CC molecule and is capable of antagonising pyrogenic exotoxin-mediated  
CC activation of Th1 lymphocytes and optionally of eliciting protective  
CC immunity against toxic shock induced by a pyrogenic exotoxin or by a  
CC mixture of at least two pyrogenic exotoxins; and (8) a method of  
CC preparing a therapeutic composition for treating a superantigen-related  
CC disorder in a mammalian subject. The substance and pharmaceutical  
CC composition have antibacterial activities, and can be used in gene  
CC therapy. The substance, peptide or CD28 molecule can be used for  
CC preparing a pharmaceutical composition for treating superantigen-related  
CC disorders caused by Staphylococcus aureus or Streptococcus pyogenes. The  
CC present sequence represents an SCD28 molecule related peptide motif,  
XX which is used in the exemplification of the present invention.  
XX Sequence 16 AA;  
Query Match 80.0%; Score 40; DB 7; Length 16;  
Best Local Similarity 90.9%; Pred. No. 9.1;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAAAM 11  
Db 1 AAAAAAAAAAAM 11  
RESULT 25

ADO43159  
ID ADO43159 standard; peptide; 30 AA.  
XX  
XX ADO43159;  
XX  
XX 29-JUL-2004 (first entry)  
XX Peptide useful for coded probe useful in biomolecule analysis.  
DE Nano-barcode; scanning probe microscopy; probe.  
XX  
XX Nano-barcode; scanning probe microscopy; probe.  
XX Synthetic.  
XX  
XX Key Location/Qualifiers  
FH Modified-site 30  
FT /note= "C-terminal amide"  
XX  
XX WO2004038037-A2.  
XX  
XX 06-MAY-2004.  
XX  
XX 22-SEP-2003; 2003WO-US029726.  
XX  
XX 20-SEP-2002; 2002US-00251152.  
XX 19-SEP-2003; 2003US-00667004.  
XX (ITLC ) INTEL CORP.  
XX  
XX Chan S, Su X, Yamakawa M;  
XX WPI; 2004-399960/37.  
XX  
XX Detecting, identifying and sequencing of biomolecules using controlled  
PT alignment of nano-barcodes encoding specific information for scanning  
PT probe microscopy, useful in the fields of molecular biology.  
XX  
XX Example 3; SEQ ID NO 1; 63pp; English.  
XX  
XX The present sequence is that of a peptide of potential use for production  
CC of a coded probe useful in the method of the invention. The sequence was  
CC obtained by solid-phase peptide synthesis, and is predicted to be an  
CC alpha helix. The invention provides methods, apparatus and compositions  
CC for the detection, identification and/or sequencing of biomolecules, such  
CC as nucleic acids or proteins. Coded probes comprising a probe molecule  
CC attached to one or more nano-barcodes are allowed to bind to target  
CC molecule(s). After binding and separation from unbound coded probes, the  
CC bound coded probes are aligned on a surface and analysed by scanning  
CC probe microscopy (SPM). The nano-barcodes may be any molecule or complex  
CC that is distinguishable by SPM, such as carbon nanotubes, fullerenes,  
CC subnanometer metallic barcodes, nanoparticle or quantum dots. The methods  
CC allow the sequencing of long nucleic acid sequences in a single  
CC sequencing run, high speed of obtaining sequence data, low cost of  
CC sequencing and high efficiency in terms of operator time, and sensitive  
CC and accurate detection and/or identification of nucleic acids with low  
CC incidence of false positive results.  
XX Sequence 30 AA;  
Query Match 80.0%; Score 40; DB 8; Length 30;  
Best Local Similarity 83.3%; Pred. No. 17;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAAAMA 12  
Db 1 AAAAAAAAAAAMA 12  
RESULT 26  
AEB28995  
ID AEB28995 standard; peptide; 30 AA.  
XX  
XX AEB28995;  
XX

DT 08-SEP-2005 (first entry)  
XX Synthetic peptide SEQ ID NO:1.  
DE nucleic acid determination; scanning probe microscopy; nanocode.  
XX  
KW Synthetic.  
XX  
OS  
XX  
FH Key Location/Qualifiers  
FT Modified-site 30  
FT /note= "CONH2"  
XX  
XX US2005147981-A1.  
XX  
XX 07-JUL-2005.  
XX  
XX 31-DEC-2003; 2003US-00750515.  
XX  
XX 31-DEC-2003; 2003US-00750515.  
XX (ITLC ) INTEL CORP.  
XX  
XX Yamakawa M, Berlin A;  
PI  
XX WPI; 2005-511772/52.  
XX  
XX Identifying target nucleic acid sequence, by contacting target nucleic acid with coded oligonucleotide probes, and identifying probes binding to target nucleic acid using scanning probe microscopy.  
XX  
XX Example 1; SEQ ID NO 1; 29pp; English.  
XX  
XX The invention relates to a method for determining a nucleotide sequence of a nucleic acid. The method comprises: (a) providing one or more coded oligonucleotide probes, each coded oligonucleotide probe comprising an oligonucleotide associated with at least one nanocode comprising a detectable non-encoding feature; (b) contacting a target nucleotide acid with the one or more coded oligonucleotide probes; and (c) identifying coded oligonucleotide probes that bind to the target nucleotide acid using scanning probe microscopy (SPM) to detect the nanocode and the detectable non-encoding feature. Nanocodes of the invention in certain aspects include detectable features beyond the arrangement of tags that encode information about the barcoded object, which assists in detecting the tags that encode information about the barcoded object. The detectable features include structures of a nanocode or associated with a nanocode, referred to as detectable feature tags, for error checking/error correction, encryption and data reduction/compression. The present sequence represents a synthetic peptide which connects four buckyballs on graphite in an example from the present invention, where the example is about the synthesis and SPM scanning of a nanocode.  
XX  
SQ Sequence 30 AA;  
Query Match 80.0%; Score 40; DB 9; Length 30;  
Best Local Similarity 83.3%; Pred. No. 17;  
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
|||||  
Db 1 AAAAAKAAAA 12  
RESULT 27  
ADN13984  
ID ADN13984 standard; protein; 92 AA.  
XX  
XX ADN13984;  
XX  
XX 29-JUL-2004 (first entry)  
DT Human prostate/colon/lung/breast cancer-related protein 1442, SEQ:1499.  
DE Human; cancer; tumour; prostate cancer; colon cancer; lung cancer;  
XX  
KW Human; cancer; tumour; prostate cancer; colon cancer; lung cancer;  
XX

KW breast cancer; drug screening; diagnosis; prognosis; prevention;  
KW gene mapping; tissue typing; tissue profiling; cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
XX WO2004039943-A2.  
XX  
XX 13-MAY-2004.  
XX  
XX 16-MAY-2003; 2003WO-US015465.  
XX  
XX 17-MAY-2002; 2002US-0381533P.  
XX  
XX 04-FEB-2003; 2003US-0445222P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Scott EM, Lamson G, Kassam A, Zhang G, Sakamoto D, Garcia PD;  
PI  
XX WPI; 2004-376173/35.  
XX  
XX N-PSDB; ADN13927.  
XX  
XX New isolated polynucleotides, useful for gene mapping or tissue typing or profiling, as diagnostic reagents, and for preventing or treating cancer, e.g. prostate, colon, or breast cancer.  
XX  
XX Claim 18; SEQ ID NO 1499; 190pp; English.  
XX  
XX The invention relates to nucleic acids (ADN12486-ADN13970) isolated from human prostate, colon, lung and breast cancer cDNA libraries, and to 57 proteins (ADN13971-ADN14027) encoded by a subset of these cDNA sequences (ADN13914-ADN13970). The invention also relates to vectors and host cells comprising a nucleic acid of the invention; a method for the recombinant production of a protein of the invention; an antibody specific for a protein of the invention; a polynucleotide library comprising at least one nucleic acid sequence of the invention; a method for detecting a cancerous cell by PCR or probe hybridisation; inhibiting a cancerous phenotype (particularly aberrant proliferation) of a cell; a method of identifying an agent that modulates the biological activity of a gene product differentially expressed in a cancerous cell compared with a normal cell; and a method of treating a cancer patient using the agent identified. The nucleic acids and polypeptides can be used to diagnose, prognose, treat or prevent cancers such as prostate, colon, lung or breast cancer, and can also be used to screen for drugs for the treatment of cancer. The nucleic acids can also be used for gene mapping, tissue typing and tissue profiling. The present sequence represents a specifically claimed cancer-related protein of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 92 AA;  
Query Match 80.0%; Score 40; DB 8; Length 92;  
Best Local Similarity 90.9%; Pred. No. 52;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 11  
|||||  
Db 20 AAAAAAAAAA 30  
RESULT 28  
ABO58367  
ID ABO58367 standard; protein; 109 AA.  
XX  
XX ABO58367;  
XX  
XX 29-JUL-2004 (first entry)  
DT Human genome derived single exon protein #4601.  
XX  
XX Human; gene expression; single exon probe; microarray;  
XX alternative splicing event; genomic alteration.  
KW

XX OS Homo sapiens.  
XX ID US2003194704-A1.  
XX AC 16-OCT-2003.  
XX DT 03-APR-2002; 2002US-00029386.  
XX DE 03-APR-2002; 2002US-00029386.  
XX KW (PENN/) PENN S G.  
XX KW (RANK/) RANK D R.  
XX KW (HANZ/) HANZEL D K.  
XX OS Penn SG, Rank DR, Hanzel DK;  
XX PN WPI; 2004-119264/12.  
XX PD  
XX PF New human genome-derived single exon nucleic acid probes useful for human  
XX PT gene expression analysis, for identifying or characterizing alternative  
XX PT splicing events, for assessing genomic alterations or as tools for  
XX PT surveying tissues.  
XX PS Claim 45; SEQ ID NO 32001; 80pp; English.  
XX CC  
XX CC The invention relates to a nucleic acid probe for measuring human gene  
XX CC expression, comprising any of the 27,400 fully defined nucleotide  
XX CC sequences in the specification, or their complements or fragments, and  
XX CC encoding at least 8 amino acids of any of the 688 amino acid sequences  
XX CC fully defined in the specification. The probe is a single exon probe that  
XX CC hybridises under high stringency conditions to a nucleic acid molecule  
XX CC expressed in human cells or tissues. Also included are a spatially-  
XX CC addressable set of single exon nucleic acid probes for measuring human  
XX CC gene expression (comprising a plurality of single exon nucleic acid  
XX CC probes cited above, where each of the plurality of probes is separately  
XX CC and addressably isolatable or amplifiable from the plurality), a single  
XX CC exon microarray for measuring human gene expression, a method of  
XX CC measuring human gene expression, a vector comprising the single exon  
XX CC probe cited above, an ORF-encoded peptide comprising at least 8  
XX CC contiguous amino acids of any of the above-mentioned amino acid  
XX CC sequences (optionally with conservative amino acid substitutions), an  
XX CC isolated antibody that binds specifically to a peptide cited above,  
XX CC methods of selling and/or licensing single exon probes or microarrays to  
XX CC a customer desiring to measure gene expression, a method of providing  
XX CC human gene expression data by subexpression, and a computer-readable  
XX CC storage medium which contains a database having a plurality of records  
XX CC (each record including data on the expression of a single exon probe  
XX CC cited above). The probe, methods and apparatus are useful in gene  
XX CC expression analysis. The probes may be used as tools for surveying  
XX CC tissues to detect the presence of expressed messages that contain their  
XX CC specific exon, or in constructing genome-derived single exon microarrays.  
XX CC In addition, the probes are used in identifying and characterising  
XX CC alternative splicing events, in detecting and characterising gross  
XX CC alterations in the genomic locus that includes their exon, in assessing  
XX CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
XX CC or in expressing the ORF-encoded peptide. The present sequence is a human  
XX CC single exon probe protein of the invention. Note: The sequence data for  
XX CC this patent did not form part of the printed specification, but was  
XX CC obtained in electronic format directly from USPTO at  
XX CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX SQ Sequence 109 AA;  
Query Match 80.0%; Score 40; DB 8; Length 109;  
Best Local Similarity 83.3%; Pred. No. 61;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
Db 36 AAAAAAAAAA 47

RESULT 29  
ABG15501  
ID ABG15501 standard; protein; 112 AA.  
XX AC ABG15501;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #15492.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS79688.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity.  
XX PS Claim 20; SEQ ID NO 45860; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain  
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX CC and in recombinant production of (II). The polynucleotides are also used  
XX CC in diagnostics as expressed sequence tags for identifying expressed  
XX CC genes. (I) is useful in gene therapy techniques to restore normal  
XX CC activity of (II) or to treat disease states involving (II). (II) is  
XX CC useful for generating antibodies against it, detecting or quantitating a  
XX CC polypeptide in tissue, as molecular weight markers and as a food  
XX CC supplement. (II) and its binding partners are useful in medical imaging  
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders  
XX CC involving aberrant protein expression or biological activity. The  
XX CC polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
XX CC amino acid sequences of the invention. Note: The sequence data for this  
XX CC patent did not appear in the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 112 AA;  
Query Match 80.0%; Score 40; DB 4; Length 112;  
Best Local Similarity 83.3%; Pred. No. 63;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
Db 83 AAAAAAAAAA 94  
RESULT 30  
AAG35185



PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 80.0%; Score 40; DB 3; Length 126;  
Best Local Similarity 90.9%; Pred. No. 71;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAAAAAAAAA 12  
| | | | | | | |  
Db 30 AAAAAAAAAA 40

RESULT 31  
ABB69576  
ID ABB69576 standard; protein; 158 AA.  
XX AC ABB69576;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 35520.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX Drosophila melanogaster.  
PN WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX N-PSDB; ABL13679.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Disclosure; SEQ ID NO 35520; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 158 AA;  
Query Match 80.0%; Score 40; DB 4; Length 158;  
Best Local Similarity 83.3%; Pred. No. 88;  
Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AAAAAAAAAA 12  
| | | | | | | |  
Db 97 AAAAAAAAAA 108  
RESULT 32  
ADU02586  
ID ADU02586 standard; protein; 191 AA.  
XX AC ADU02586;  
XX DT 27-JAN-2005 (first entry)  
XX DE Novel human polypeptide seqid 1053.  
XX cytosolic; antipsoriatic; antiinflammatory; gene therapy; Nanodisc;  
KW proliferative disorder; inflammatory disorder; immune disorder;  
KW metabolic disorder; bone disorder; CNS disorder; cancer; psoriasis;  
KW ulcerative colitis; human.  
XX Homo sapiens.  
XX WO2004093804-A2.  
XX 04-NOV-2004.  
XX 19-APR-2004; 2004WO-US012047.

PR 18-APR-2003; 2003US-0463708P.  
PR 18-APR-2003; 2003US-0463732P.  
PR 02-MAY-2003; 2003US-0467199P.  
PR 02-MAY-2003; 2003US-0467230P.  
PR 19-MAY-2003; 2003US-0471306P.  
PR 19-MAY-2003; 2003US-0471336P.  
PR 08-JUL-2003; 2003US-0485223P.  
PR 08-JUL-2003; 2003US-0485224P.  
PR 14-JUL-2003; 2003US-0486446P.  
PR 14-JUL-2003; 2003US-0486480P.  
PR 08-AUG-2003; 2003US-0493573P.  
PR 08-AUG-2003; 2003US-0493577P.  
PR 08-SEP-2003; 2003US-0505059P.  
XX (FIVE-) FIVE PRIME THERAPEUTICS INC.  
XX  
XX Lee E, Hestir K, Chu K, Masuoka L, Williams LT;  
XX WPI; 2004-775861/76.  
XX N-PSDB; ADU01854.  
XX  
XX New first nucleic acid molecule comprising a polynucleotide sequence  
PT given in the specification, useful in preparing a composition for  
PT diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.  
XX  
XX Claim 14; SEQ ID NO 1053; 291pp; English.  
XX  
XX The invention describes a new first nucleic acid molecule comprising a  
CC polynucleotide sequence given in the specification. Also described are:  
CC an animal injected with the nucleic acid molecule; a second nucleic acid  
CC molecule comprising a second polynucleotide sequence that is at least  
CC about 70, 80, 90 or 95% homologous to the first nucleic acid molecule or  
CC that hybridises to the first polynucleotide sequence under high  
CC stringency conditions; a vector comprising the nucleic acid molecule; a  
CC promoter that drives the expression of the nucleic acid molecule; a  
CC host cell transformed, transfected, transduced or infected with the  
CC nucleic acid molecule; a nucleic acid composition comprising a carrier or  
CC a buffer and one or more compositions comprising the nucleic acid  
CC molecule, vector or host cell; a substantially purified polypeptide; an  
CC animal injected with the polypeptide; a polypeptide composition  
CC comprising the polypeptide molecule and a carrier or buffer; a cell  
CC culture medium comprising the polypeptide or transfected cells  
CC transfected with the polynucleotide; making a transformed, transfected,  
CC transduced, or infected host cell; synthesising Nanodiscs simultaneously  
CC and for synthesising a series of simultaneously-synthesised Nanodiscs  
CC sequentially utilising a dynamic system; preparing a hydrophobic protein  
CC for determination of crystal structure; immunising a non-human animal;  
CC screening for modulators of hydrophobic protein activity; a diagnostic  
CC kit; determining the presence of the nucleic acid molecule or its  
CC complement; determining the presence of an antibody to the polypeptide in  
CC a sample; an antibody specifically recognising, binding to or modulating  
CC the biological activity of at least one polypeptide encoded by a nucleic  
CC acid molecule or its biologically active fragment; an antibody  
CC composition comprising the antibody and a carrier; a bacteriophage, where  
CC the antibody is displayed on the bacteriophage; a bacterial cell  
CC comprising the bacteriophage; a non-human animal injected with the  
CC antibody composition; a host cell that secretes the antibody; making an  
CC antibody; diagnosing a disease, disorder, syndrome, or condition  
CC comprising cancer, or proliferative, inflammatory, immune, metabolic,  
CC bone, CNS, genetic, bacterial and viral diseases, disorders, syndromes or  
CC conditions in a patient; a modulator composition comprising a modulator  
CC and a carrier; gene therapy; prophylactic or therapeutic treatment of a  
CC subject; an isolated modified cell comprising at least one first  
CC heterologous nucleic acid molecule, where the first heterologous nucleic  
CC acid molecule comprises a first polynucleotide sequence that encodes a  
CC first polypeptide; a non-human animal deficient in the polypeptide or  
CC that over-expresses the polypeptide; isolated tissues derived from the  
CC non-human animal; and one or more cells derived from the non-human  
CC animal. The nucleic acid is useful in preparing a composition for  
CC diagnosing or treating e.g., cancer, psoriasis or ulcerative colitis.  
CC This is the amino acid sequence of a novel human polypeptide of the  
CC invention.  
XX

SQ Sequence 191 AA;  
Query Match 80.0%; Score 40; DB 8; Length 191;  
Best Local Similarity 90.9%; Pred. No. 1,le-02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAM 11  
|||  
DB 20 AAAAAAAAAAAM 30  
|||  
RESULT 33  
ADT57694  
ID ADT57694 standard; protein; 227 AA.  
XX  
XX AC ADT57694;  
XX  
XX DT 13-JAN-2005 (first entry)  
XX  
XX DE Plant polypeptide, SEQ ID 7771.  
XX  
XX KW Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
XX OS Viridiplantae.  
XX  
XX PN US2004216190-A1.  
XX  
XX PD 28-OCT-2004.  
XX  
XX PF 18-DEC-2003; 2003US-00739930.  
XX  
XX PR 28-APR-2003; 2003US-00424599.  
XX  
XX PR 28-APR-2003; 2003US-00425115.  
XX  
XX PA (KOVA/) KOVALIC D K.  
XX  
XX PI Kovalic DK;  
XX  
XX WPI; 2004-757369/74.  
XX  
XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX  
XX Claim 2; SEQ ID NO 7771; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and

CC development under at least one stress condition. The polynucleotide may  
 CC also encode a plant transcription factor. The methods and compositions of  
 CC the present invention are useful in the field of biochemistry and  
 CC genetics, in particular for producing transgenic plants with improved  
 CC biological characteristics such as increased yield, improved nitrogen  
 CC flow, increasing plant tolerance to cold or heat, improving plant  
 CC tolerance to extreme osmotic and drought conditions, and improving plant  
 CC tolerance to plant pests or pathogens. They can also be used in physical  
 CC arrays of molecules, plant breeding markers, computer-based storage and  
 CC analysis systems. The present sequence is one of the 5544 plant protein  
 CC sequences of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20040216190.

XX SQ Sequence 227 AA;

Query Match 80.0%; Score 40; DB 8; Length 227;  
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
 Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 12  
 ||| |||||  
 Db 14 AAAAAAAAAA 25

RESULT 34

ADH75874  
 ID ADH75874 standard; protein; 231 AA.

XX AC ADH75874;

XX DT 15-APR-2004 (first entry)

XX DE Corn C-repeat/DRE binding factor (CBF) G3441.

XX KW Plant; C-repeat/DRE binding factor; CBF; transgenic; biomass;  
 KW cold inducible promoter; environmental stress; cold tolerance;  
 KW cell injury; growth inhibition; drought; high salt condition.

XX OS Zea mays.

XX PN US2003233680-A1.

XX PD 18-DEC-2003.

XX PF 23-APR-2003; 2003US-00421138.

XX PR 04-SEP-1996; 96US-00706270.

XX PR 03-FEB-1998; 98US-00017575.

XX PR 03-FEB-1998; 98US-00017816.

XX PR 03-FEB-1998; 98US-00018227.

XX PR 03-FEB-1998; 98US-00018233.

XX PR 03-FEB-1998; 98US-00018234.

XX PR 03-FEB-1998; 98US-00018235.

XX PR 23-NOV-1998; 98US-00198119.

XX PR 28-JAN-1999; 99WO-US001895.

XX PR 17-NOV-1999; 99US-0166228P.

XX PR 17-APR-2000; 2000US-0197899P.

XX PR 26-MAY-2000; 2000US-00580377.

XX PR 22-AUG-2000; 2000US-0227439P.

XX PR 15-SEP-2000; 2000US-00601802.

XX PR 16-NOV-2000; 2000US-00713994.

XX PR 01-FEB-2001; 2001US-00773990.

XX PR 26-NOV-2001; 2001US-00996140.

XX (THOM/) THOMASHOW M.

PA (STOC/) STOCKINGER E J.

PA (JAGL/) JAGLO K.

PA (GILM/) GILMOUR S J.

PA (ZARK/) ZARKA D.

PA (JIAN/) JIANG C.

PA (ZHAN/) ZHANG J.

PA (HAAK/) HAAKE V.  
 PA (RIEC/) RIECHMANN J L.  
 PA (SHER/) SHERMAN B K.  
 PA (FROM/) FROMM M.  
 PA (CENT/) CENTURY K.

PI Thomashow M, Stockinger EJ, Jaglo K, Gilmour SJ, Zarka D;  
 PI Jiang C, Zhang J, Haake V, Riechmann JL, Sherman BK, Fromm M;  
 PI Century K;

XX WPI; 2004-070474/07.

XX PT New transgenic plants comprising C-repeat/DRE binding factor, useful in  
 PT producing phenotypes with increased tolerance to an environmental stress,  
 PT e.g. drought, cold, freezing or high salt.

XX PS Example 20; SEQ ID NO 314; 223pp; English.

XX CC The invention relates to a transgenic plant having an altered trait as  
 CC compared to a wild type or untransformed plant. The trait is altered,  
 CC i.e. increasing or decreasing levels of at least one cell protectant in  
 CC cells of the transgenic plant or altered, i.e. increasing or decreasing  
 CC levels of biomass, where the transgenic plant is transformed with a  
 CC recombinant polynucleotide comprising a nucleotide sequence that encodes  
 CC a C-repeat/DRE binding factor (CBF) comprising an AP2 domain. Also  
 CC included are a cold inducible promoter comprising a Arabidopsis thaliana  
 CC CBPI under the control of the Rabi8 or Dreb2a promoters, a method of  
 CC identifying the levels of a cell protectant in a cell, a method for  
 CC modifying the levels of a cell protectant in a plant cell, a method for  
 CC improving the tolerance of a plant cell to an environmental stress, a  
 CC method of screening for a CBF, a method for identifying a CBF, a method  
 CC of enhancing cold tolerance and a method of altering the biomass of a  
 CC plant. The improved environmental stress is selected from a decrease in  
 CC the extent of a plant's or cell's injury, plant's growth inhibition,  
 CC growth inhibition, increase in survival rate after exposure to cold  
 CC temperatures, survival rate after exposure to freezing temperatures,  
 CC extent of survival of a plant after exposure to drought conditions,  
 CC extent of survival of a cell after exposure to drought conditions,  
 CC survival of a plant after exposure to high salt conditions, and extent of  
 CC survival of a cell after exposure to high salt conditions. The plants are  
 CC useful for further production of plants with better quality and which has  
 CC an increased tolerance to an environmental stress, e.g. drought, cold,  
 CC freezing or high salt. The present sequence represents a plant CBF or  
 CC homologue.

XX SQ Sequence 231 AA;

Query Match 80.0%; Score 40; DB 8; Length 231;

Best Local Similarity 90.9%; Pred. No. 1.3e+02;

Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 11  
 ||| |||||  
 Db 170 AAAAAAAAAA 180

RESULT 35

ADI42782

ID ADI42782 standard; protein; 231 AA.

XX AC ADI42782;

XX DT 22-APR-2004 (first entry)

XX DE Plant transcription factor #450.

XX KW transgenic; plant; enhanced tolerance to abiotic stress;

KW glyophosphate tolerance; hormone sensitivity; disease resistance;

KW sugar sensing; flowering; flower structure; stem bifurcation;

KW branching pattern; apical dominance; trichome; stem morphology;

KW root growth; root hair; seed development; cell proliferation;

KW cell differentiation; premature senescence; necrosis; plant size;

KW leaf morphology; seed morphology; seed biochemistry; root anthocyanin;

KW plant anthocyanin; light response; shade avoidance; bioinformatic;  
KW transcription factor; gene; ds.  
XX

OS Zea mays.

XX US2004019927-A1.

XX 29-JAN-2004.

XX 25-FEB-2003; 2003US-00374780.

XX 18-APR-2001; 2001US-00837944.

XX (SHER/) SHERMAN B K.

XX (RIEC/) RIECHMANN J L.

XX (JIAN/) JIANG C.

XX (HEAR/) HEARD J E.

XX (HAAR/) HAAKE V.

XX (CREE/) CREELMAN R A.

XX (RATC/) RATCLIFFE O.

XX (ADAM/) ADAM L J.

XX (REUB/) REUBER T L.

XX (KEDD/) KEDDIE J.

XX (BROU/) BROWN P E.

XX (FILG/) FILGRIM M L.

XX (DUBE/) DUBELL A N.

XX (PINE/) PINEDA O.

XX (YUGG/) YU G.

XX Sherman BK, Riechmann JL, Jiang C, Heard JE, Haake V;

PI Creelman RA, Ratcliffe O, Adam LJ, Reuber TL, Keddie J, Brown PE;

PI Pilgrim ML, Dubell AN, Pineda O, Yu G;

XX WPI; 2004-132245/13.

XX New transgenic plant comprising a recombinant polynucleotide of any one  
PT of more than 500 nucleotide sequences, useful in bioinformatic search  
PT methods.  
XX

PS Claim 1; SEQ ID NO 1245; 435pp; English.

XX The invention describes a transgenic plant comprising a recombinant  
CC polynucleotide of any one of more than 500 nucleotide sequences fully  
CC defined in the specification or its complement. The method of the  
CC invention can be used to produced a plant having altered traits such as:  
CC enhanced tolerance to abiotic stress; glyphosphate tolerance; hormone  
CC sensitivity; disease resistance; sugar sensing; early or late flowering;  
CC altered flower structure; change in stem bifurcations, altered branching  
CC pattern, reduced apical dominance, reduced trichome density; lack of  
CC trichomes; reduced ectopic trichome development; altered trichome  
CC development; increase in trichome number; altered stem morphology;  
CC increased root growth; increased root hairs; altered seed development;  
CC altered cell proliferation or cell differentiation; rapid development;  
CC premature senescence; increased necrosis; increase in seedling or plant  
CC size; decreased plant size; leaf morphology; seed morphology; seed  
CC biochemistry; increase in root anthocyanins; increase in plant  
CC anthocyanins; or alteration in light response or shade avoidance. The  
CC transgenic plant, polynucleotides and polypeptides are useful in  
CC bioinformatic search methods. This is the amino acid sequence of a plant  
CC transcription factor, and an orthologue of Arabidopsis thaliana  
CC transcription factors isolated in the invention, that can be used in the  
CC creation of a transgenic plant with altered traits.  
XX

XX Sequence 231 AA;

Query Match 80.0%; Score 40; DB 8; Length 231;  
Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 11

||| |||||

Db 170 AAAAAAAAAA 180

RESULT 36

AEE27925  
ID AEE27925 standard; protein; 231 AA.

XX AEE27925;

XX 09-FEB-2006 (first entry)

XX CBF clade transcription factor G3441 SEQ ID NO 178.

XX transformation; transgenic plant; plant; abiotic stress tolerance;  
KW transcription factor; CBF.

XX Zea mays.

XX WO2005112608-A2.

XX 01-DEC-2005.

XX 19-MAY-2005; 2005WO-US017583.

XX 21-MAY-2004; 2004US-0573443P.

XX (UNMS ) UNIV MICHIGAN STATE.

XX Thomasow M, Gilmour SJ, Cook DP, Canella D;

XX WPI; 2006-010542/01.

XX N-PSDB; AEE27924.

XX Producing transformed plant, by transforming a target plant with a first  
PT polynucleotide, which is a member of the CBF clade of transcription  
PT factors, and which is mutagenized, and identifying a transformed plant.

XX Example 9; SEQ ID NO 178; 225pp; English.

XX The invention describes a method of producing a transformed plant  
CC comprises transforming a target plant with the first polynucleotide,  
CC which is a member of the CBF clade of transcription factors, and is  
CC mutagenized to produce a second or a third polynucleotide, where the  
CC second or third polypeptide has an amino acid sequence different than the  
CC first polypeptide, and identifying a transformed plant comprising the  
CC second polynucleotide or the third polynucleotide. Also described are: a  
CC transgenic plant produced by the method above; a seed produced by the  
CC transgenic plant, where the seed comprises the second or third  
CC polynucleotide; and a transgenic plant comprising a recombinant,  
CC overexpressed mutated polynucleotide that encodes a mutant member of the  
CC CBF clade of transcription factor polypeptides, where the transgenic  
CC plant exhibits fewer or reduced adverse morphological or developmental  
CC effects than a plant that overexpresses a wild-type form of the mutated  
CC polynucleotide. The method is useful for producing a transformed plant  
CC having increased abiotic stress tolerance, as compared to wild type or  
CC control plants. This is the amino acid sequence of a CBF clade  
CC transcription factor.

XX Sequence 231 AA;

Query Match 80.0%; Score 40; DB 10; Length 231;  
Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAAAAAAAAA 11

||| |||||

Db 170 AAAAAAAAAA 180

RESULT 37

AEE27977  
ID AEE27977 standard; protein; 231 AA.

XX AEE27977;



DT 09-FEB-2006 (first entry)  
XX CBF clade transcription factor G3441.  
DE transformation; transgenic plant; plant; abiotic stress tolerance;  
XX transcription factor; CBF.  
XX Unidentified.  
OS  
PN WO2005112608-A2.  
XX  
XX 01-DEC-2005.  
XX  
XX 19-MAY-2005; 2005WO-US017583.  
XX  
XX 21-MAY-2004; 2004US-0573443P.  
XX (UNMS ) UNIV MICHIGAN STATE.  
XX  
XX Thomasow M, Gilmour SJ, Cook DD, Canella D;  
XX WPI; 2006-010542/01.  
XX  
XX Producing transformed plant, by transforming a target plant with a first  
PT polynucleotide, which is a member of the CBF clade of transcription  
PT factors, and which is mutagenized, and identifying a transformed plant.  
XX  
XX Disclosure; Fig 3; 225pp; English.  
XX  
XX The invention describes a method of producing a transformed plant  
CC comprises transforming a target plant with the first polynucleotide,  
CC which is a member of the CBF clade of transcription factors, and is  
CC mutagenized to produce a second or a third polynucleotide, where the  
CC second or third polypeptide has an amino acid sequence different than the  
CC first polypeptide, and identifying a transformed plant comprising the  
CC second polynucleotide or the third polynucleotide. Also described are: a  
CC transgenic plant produced by the method above; a seed produced by the  
CC transgenic plant, where the seed comprises the second or third  
CC polynucleotide; and a transgenic plant comprising a recombinant,  
CC overexpressed mutated polynucleotide that encodes a mutant member of the  
CC CBF clade of transcription factor polypeptides, where the transgenic  
CC plant exhibits fewer or reduced adverse morphological or developmental  
CC effects than a plant that overexpresses a wild-type form of the mutated  
CC polynucleotide. The method is useful for producing a transformed plant  
CC having increased abiotic stress tolerance, as compared to wild type or  
CC control plants. This is the amino acid sequence of a CBF clade  
XX transcription factor.  
XX  
SQ Sequence 231 AA;  
Query Match 80.0%; Score 40; DB 10; Length 231;  
Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAM 11  
|||  
DB 170 AAAAAAAAAAAM 180  
RESULT 38  
ADE08006  
ID ADE08006 standard; protein; 232 AA.  
XX  
AC ADE08006;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
XX Novel protein (useful for identifying genetic disorders) #161.  
XX novel gene; novel protein; tissue marker; molecular weight marker;  
KW chromosome marker; genetic disorder.  
XX  
OS Unidentified.

XX WO2003054152-A2.  
XX  
XX 03-JUL-2003.  
XX  
XX 10-DEC-2002; 2002WO-US039555.  
XX  
XX 10-DEC-2001; 2001US-0339739P.  
PR 11-DEC-2001; 2001US-0339453P.  
PR 14-MAR-2002; 2002US-0365091P.  
PR 14-MAR-2002; 2002US-0365384P.  
PR 12-APR-2002; 2002US-037281P.  
PR 12-APR-2002; 2002US-037281P.  
PR 22-APR-2002; 2002US-00128558.  
PR 24-APR-2002; 2002US-0376045P.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;  
XX  
DR WPI; 2003-569235/53.  
DR N-PSDB; ADE07095.  
XX  
XX New polynucleotides, useful for expressing recombinant proteins for  
PT analysis, characterization or therapeutic use, or as markers for tissues  
PT in which the corresponding protein is preferentially expressed.  
XX  
XX Claim 20; SEQ ID NO 1072; 1177pp; English.  
XX  
XX The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present amino acid sequence represents a protein  
CC of the invention.  
XX  
SQ Sequence 232 AA;  
Query Match 80.0%; Score 40; DB 7; Length 232;  
Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AAAAAAAAAAAM 11  
|||  
DB 5 AAAAAAAAAAAM 15  
RESULT 39  
ADU40376  
ID ADU40376 standard; protein; 232 AA.  
XX  
AC ADU40376;  
XX  
XX 27-JAN-2005 (first entry)  
XX  
XX Novel human polypeptide seq id 161.  
XX  
XX cytostatic; antianemic; immunosuppressive; neuroprotective;  
KW antirheumatic; antiarthritic; muscular; osteopathic; vulnary;  
KW antiulcer; antiinflammatory; CNS; nootropic; antiparkinsonian;  
KW anticonvulsant; cerebroprotective; vasotropic; gene therapy;  
KW gene mapping; forensic; mutation identification; biodiversity; cancer;  
KW haematopoiesis; myeloid cell disorder; lymphoid cell disorder;  
KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
KW myasthenia gravis; tissue growth; tissue regeneration; tissue repair;  
KW tissue replacement; burn; incision; ulcer; osteoporosis; osteoarthritis;  
KW bone degenerative disorder; periodontal disease; nervous system disease;  
KW neuropathy; mechanical disorder; traumatic disorder; nerve injury;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease;

XX amytrophic lateral sclerosis; Shy-Drager syndrome; stroke; human.  
XX Homo sapiens.  
XX US2004219521-A1.  
XX  
XX PD 04-NOV-2004.  
XX PF 22-APR-2002; 2002US-00128558.  
XX  
XX PR 21-JAN-2000; 2000US-00488725.  
XX PR 25-JAN-2000; 2000US-00494404.  
XX PR 25-APR-2000; 2000US-00552317.  
XX PR 22-DEC-2000; 2000WO-US035017.  
XX PR 25-JAN-2001; 2001WO-US002623.  
XX PR 05-FEB-2001; 2001WO-US003800.  
XX PR 26-FEB-2001; 2001WO-US004927.  
XX PR 05-MAR-2001; 2001WO-US004941.  
XX PR 30-MAR-2001; 2001WO-US008631.  
XX PR 18-APR-2001; 2001WO-US008656.  
XX PR 11-DEC-2001; 2001US-0339453P.  
XX (TANG/) TANG Y T.  
XX PA (WANG/) WANG Z.  
XX PA (WENG/) WENG G.  
XX PA (BOYL/) BOYLE B J.  
XX PA (DRMA/) DRMANAC R T.  
XX  
XX PI Tang YT, Wang Z, Weng G, Boyle BJ, Drmanac RT;  
XX  
XX WPI; 2005-010094/01.  
XX N-PSDB; ADU40252.  
XX  
XX PR New isolated polynucleotide, useful in therapeutic (i.e. for treating  
XX e.g., cancer, neurodegenerative disorders), diagnostic (as expressed  
XX sequence tags for identifying expressed genes) and research methods  
XX (e.g., gene mapping).  
XX  
XX Claim 10; SEQ ID NO 161; 138pp; English.  
XX  
XX The invention describes an isolated polynucleotide comprising a  
XX nucleotide sequence selected from 124 nucleotide sequences (SEQ ID NO: 1-  
XX 124), or a polynucleotide encoding a polypeptide with biological  
XX activity, where the polynucleotide hybridises to the nucleotide sequence  
XX under stringent hybridisation conditions or has greater than 99% sequence  
XX identity with the nucleotide sequence. The polynucleotide or the encoded  
XX polypeptide is useful in therapeutic, diagnostic (as expressed sequence  
XX tags for identifying expressed genes) and research methods such as gene  
XX mapping, forensics, identification of mutations responsible for genetic  
XX disorders or other traits, to assess biodiversity, and to produce many  
XX other types of data and products dependent on DNA and amino acid  
XX sequences. The polynucleotide or the polypeptide is useful in methods for  
XX preventing and/or treating disorders involving aberrant protein  
XX expression or biological activity e.g., cancer, as nutritional sources or  
XX supplements. The polypeptide is useful for regulating hematopoiesis and  
XX thus treating myeloid or lymphoid cell disorders and for treating  
XX autoimmune disorders such as multiple sclerosis, rheumatoid arthritis,  
XX and myasthenia gravis, in bone, cartilage, tendon, ligament and/or nerve  
XX tissue growth or regeneration, as well as in wound healing and tissue  
XX repair and replacement, and in healing of burns, incisions and ulcers,  
XX for treating osteoporosis, osteoarthritis, bone degenerative disorders,  
XX or periodontal disease. The composition is useful in assays for  
XX proliferation and differentiation of various hematopoietic lines, assays  
XX for embryonic stem cell differentiation, for proliferating neural cells  
XX and for regenerating nerve and brain tissue i.e., for treating central  
XX and peripheral nervous system diseases and neuropathies, as well as  
XX mechanical and traumatic disorders, which involve degeneration, death or  
XX trauma to neural cells or nerve tissue, particularly diseases such as  
XX peripheral nerve injuries, peripheral neuropathy and localised  
XX neuropathies, Alzheimer's disease, Parkinson's disease, Huntington's  
XX disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome, as well  
XX as stroke. This sequence represents a novel polypeptide of the invention.

## SQ Sequence 232 AA;

Query Match 80.0%; Score 40; DB 9; Length 232;  
Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAM 11  
||| |||||  
Db 5 AAAAAAAAAAAM 15

## RESULT 40

ADV70174

ID ADV70174 standard; protein; 236 AA.

AC ADV70174;

DT 10-MAR-2005 (first entry)

XX Tumor-associated antigenic target polypeptide TAT285.

DE cytostatic; diagnosis; therapy; tumor;

KW tumor-associated antigenic target polypeptide; TAT.

XX Homo sapiens.

XX WO2004112829-A2.

XX 29-DEC-2004.

XX 21-MAY-2004; 2004WO-US016121.

XX 23-MAY-2003; 2003US-0473238P.

XX 27-FEB-2004; 2004US-0548299P.

XX (GETH ) GENENTECH INC.

XX Phillips H;

XX WPI; 2005-048766/05.

XX N-PSDB; ADV70078.

XX Treating a mammal having a tumor of glial origin comprising cells that  
XX express a type A or B glial tumor antigen by contacting the cells with a  
XX composition comprising first and second binding agents.  
XX Disclosure; SEQ ID NO 98; 374pp; English.  
XX The invention describes a method of treating a mammal having a tumor of  
XX glial origin comprising cells that express a type A or B glial tumor  
XX antigen comprises contacting the cells with a composition of: a first  
XX binding agent comprising a first antibody, oligopeptide or organic  
XX molecule that binds to a type A or B glial tumor antigen; and a second  
XX binding agent comprising a second antibody, oligopeptide or organic  
XX molecule that binds to a type B or A glial tumor antigen. Also described  
XX is a method of determining the presence of a type A or B glial tumor in a  
XX mammal. The method is useful in treating a mammal having a tumor of glial  
XX origin comprising cells that express a type A or B glial tumor antigen.  
XX This sequence represents a human tumor-associated antigenic target  
XX polypeptide.

## SQ Sequence 236 AA;

Query Match 80.0%; Score 40; DB 9; Length 236;  
Best Local Similarity 90.9%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAAAAAAAAAAM 11  
||| |||||  
Db 5 AAAAAAAAAAAM 15

Search completed: September 9, 2006, 22:48:14

Job time : 92.924 secs